

2

DEVELOPING RESEARCH QUESTIONS AND HYPOTHESES IN MIXED METHODS RESEARCH

One of the most important and often challenging steps of the research process is specifying novel, answerable, and interesting research questions.¹ Good questions (1) address gaps or problematic assumptions in a field, (2) are rooted in theory—or at least a notion of what ideas need further refinement, and (3) suggest clear empirical possibilities, whether exploratory or explanatory (Alford 1998; Timmermans and Tavory 2012). When researchers are prepared to predict the direction and nature of the results of an analysis, they develop hypotheses. Hypotheses are typically only used prior to the analysis of quantitative data, although they could also be appropriate for a confirmatory analysis of qualitative data (Hesse-Biber and Dupuis 2000). In other words, when researchers have theoretically based propositions to test with their data, they may articulate hypotheses prior to data analysis. Hypotheses and/or research questions are vital to any research project, including mixed methods research studies. Further, in mixed methods research, one strand of research can assist in the development of research questions and hypotheses for the other.

There are two general ways of conceptualizing how research questions and hypotheses relate to a project. The most common view is that they are an essential starting point and that methods of inquiry are selected based on the research questions and hypotheses of a study and not vice versa (Greene 2007; Johnson and Onwuegbuzie 2004; Newman and Ridenour 2008;

¹ Research questions are very similar to what some, especially those who write proposals for National Institutes of Health grants, might call aims. Aims are statements of what the research project will accomplish, and research questions are literal questions to be answered by the research project.

Onwuegbuzie and Leech 2006; Plano Clark and Badiie 2010; Tashakkori and Teddlie 2009). This implies one must start with clearly defined research questions and hypotheses, and the rest of the study is developed based on those. This is certainly the ideal, but increasingly, scholars are openly embracing the complex and dynamic nature of research questions and hypotheses and how they evolve in practice (Maxwell and Loomis 2003).

Although one's research design, methods, data, and analyses must be a good fit for answering the research question(s) or testing the hypotheses, research questions and hypotheses do not develop free of other influences. They are shaped by training, fields, prior experience with certain methods, positionality, and expertise (Maxwell and Loomis 2003; Pearce 2015). As a result, research questions and hypotheses can both shape and be shaped by the research design, the data available, and findings along the way (Maxwell and Loomis 2003; Plano Clark and Badiie 2010). Research questions and hypotheses are also influenced by researchers' interests and biases, which are both informed by their own experiences (Jacobson and Mustafa 2019). The process of developing research questions and designing a study requires attention to some elements of social life to the exclusion of others, and those choices are not value free.

Whether fully antecedent or emerging and evolving, research questions and/or hypotheses are essential for focusing the project and serve as a link between theory and data. We thus divide this chapter into two main sections, first focusing on the case in which a set of research questions or hypotheses are fully determined at the start, discussing how this works in both concurrent and sequential designs, sharing different forms research questions, hypotheses, and/or aims might take. The second part of the chapter explains how research questions or hypotheses might emerge across the life of a project in two particular ways: (1) using qualitative data analysis to build theory that guides research question and hypothesis development for a new strand of research and (2) using results from quantitative data analysis to guide the development of research questions that lead to further understanding of the processes that explain them. We also discuss the challenges of writing grant proposals that include emergent strands of research.

PREDETERMINED RESEARCH QUESTIONS OR HYPOTHESES FOR MIXED METHODS RESEARCH

When proposing and planning a mixed methods research project, a set of good research questions is essential. Some hypotheses are likely to follow for any particularly deductive or confirmatory strands of research, but starting

with the research questions, there are a few general approaches that one could use: a broad research question that guides all the strands of research, separate research questions for each strand established at the beginning of a project, or separate research questions in which one or more evolve from the analysis of one type of data. The choice of approach should be based on the temporal order in which the analyses unfold, the balance of emphasis on the different strands of research, and the integration of methods being used.

One Overarching, Preestablished Question

One option is to use a single, broad research question that guides all strands of a research project (Tashakkori and Creswell 2007). In the beginning of one of our own projects (Krull, Pearce, and Jennings 2021), concerned by the possibility that prior research findings have largely been driven by the number of white and middle-class participants in nationally representative samples, we asked,

How are key dimensions of religiosity (biblical literalism, personal religiosity, and religious service attendance) differentially related to young women's reproductive and contraceptive knowledge by social class and race?

This was a concurrent mixed methods project, equally balancing the independent analyses of quantitative survey data and qualitative interview data. Using Relationship Dynamics and Social Life (Barber, Kusunoki, and Gatny 2011) survey data, we endeavored to test whether correlations between young women's religiosity and their reproductive and contraceptive knowledge varied, in statistically significant ways, by racial identification and level of parental education. The analysis of the qualitative interview data, from the National Study of Youth and Religion (Smith and Denton 2009), focused on how young women with different religious, racial, and social class identities spoke about their sexual and reproductive knowledge and experiences. We focused on if and how discourse regarding sex and contraception varied by subgroup or not. Both sets of findings were then integrated to richly describe subgroup differences in reproductive and sexual knowledge.

Two elements of the Krull, Pearce, and Jennings (2021) study were conducive to using a single, overarching research question to guide the research. First, there was heavy overlap between the survey and interview questions asked of respondents, which allowed for a focus on the same themes across data types. Second, the data were analyzed concurrently, and the results were presented together, such that the research questions guiding these analyses

had to be established from the outset. This single/common/shared research question(s) approach is ideal for research projects in which complementary information about the same topic is available in different data formats, with each type of data providing a different perspective on the same question(s).

Separate Strand-Specific Questions

Another approach to designating research questions in a mixed methods project is to have separate questions for the different strands of research. This is most useful when the strands of research unfold in a sequential manner. Two questions can also be used to signal an unequal focus on (or balance between) the strands of research or when there is less integration across strands. An example of a separate questions approach can be found in an article by Cowan et al. (2022) about the ways in which abortion attitudes may or may not align with a willingness to help friends or family members who are planning to have an abortion. They write their two research questions as follows:

First, how are abortion attitudes associated with people's willingness to offer different kinds of help to friends and family members pursuing an abortion? Second, among individuals who are willing to help in the face of conflicting abortion attitudes, how do they explain their discordant benevolence? (4)

This is an example of a mixed methods study in which the first question is designed to establish a pattern or association using quantitative data. In this case, the goal was identifying the extent to which "discordant benevolence" exists (Cowan et al. 2022:1). In other words, how common is it that people are opposed to abortion being legally available and yet are willing to offer emotional and/or material support to friends or family members who decide to have an abortion? This is a relatively straightforward question answered by analyses of survey data on abortion attitudes.

The second research question is distinct from the first. It proceeds to further reveal what underlies an experience of discordant benevolence. As the authors convincingly argue, the best approach for examining how people make sense of the seeming contradiction in their attitudes and the behaviors they are willing to employ is the analysis of qualitative interview data in which participants talk through their willingness to embrace discordant benevolence (Cowan et al. 2022:6). These two questions, one after the other, clearly imply a sequential analysis of two types of data.

Of course, any two questions can also be hybridized into one, much like the question reflected in Cowan et al.'s (2022) title: how and why do people

help others in the face of conflicting values? This still implies two steps in the research process. In addition, although we have thus far focused on conceptual or theoretical questions that suggest empirical avenues of analysis, research questions or aims can also be methodologically focused (Plano Clark and Badiee 2010). In Cowan et al.'s (2022) case, that would mean wording the questions as follows: (1) What percentage of General Social Survey respondents indicate that they are both opposed to abortion and would help a friend or family member who is pursuing an abortion? (2) What do semi-structured interviews reveal about the ways that individuals understand their discordant benevolence? Methodological details are mentioned explicitly as part of the questions.

For those working with one very general, hybrid, or integrated research question to guide a convergent, concurrent design, remnants of the incommensurability perspective² often surface, challenging the notion that qualitative and quantitative data analyses could be used in the service of answering a common research question. In other words, if there is a general overarching or hybrid question for which the analysis of qualitative data is going to be compared to the analysis of quantitative data, some might still argue that the guiding philosophies of the qualitative and quantitative paradigms of research are too different to compare and contrast the results in a meaningful way. Our position is more in line with the sentiments of Howard S. Becker (1996), who said,

Both kinds of research try to see how society works, to describe social reality, to answer specific questions about specific instances of social reality. Some social scientists are interested in very general descriptions, in the form of laws about whole classes of phenomena. Others are more interested in understanding specific cases, how those general statements worked out in this case. But there's a lot of overlap. (53)

As long as mixed methods researchers are cognizant of the guiding philosophies and strengths and weaknesses of a given approach, they can use what may sometimes seem to be conflicting or incommensurable results to their advantage in building rich layers of evidence regarding the social dynamics under study (Pearce 2015).³

² The incommensurability thesis argues that quantitative and qualitative research are based in two distinct paradigms that are incompatible and therefore cannot be mixed (Guba and Lincoln 1994).

³ For discussion of the philosophical foundations of mixed methods research, see Shan (2022).

Prestablished Research Aims in Proposal Writing

Developing one or more research questions from the beginning of a project is also essential for constructing a research proposal. Funders want to know what the expected payoff of the research is going to be. To this end, researchers must have a clear sense of their research questions and then convert these questions into aims. Research aims are a lot like questions, conveying the goals of a project, but are framed as statements. Research aims are a requirement of research proposals to the National Institutes of Health (NIH), so to demonstrate a more methods focused set of aims, we draw on a funded R21 proposal for which one of us (Pearce) is the principal investigator and present the first research aim.

Aim 1: Collect family network data, health behavior and psychosocial measures, and maternal health outcome data for 400 pregnant people sampled from electronic medical records (EMRs) from the predominant health care system in two demographically diverse counties in North Carolina. First, in face-to-face interviews, using social network software on tablets, participants will name all significant, self-defined family members (related or not), with an emphasis on both supportive and strained ties. Then, participants will describe their family network members and characterize all ties within the network in terms of social support/conflict, social engagement, and social influence. We will also ask open-ended questions to learn why some family members are unreported or to discover salient aspects of family relations not measured by our structured questions. In addition, we will use well-validated scales to measure psychosocial well-being, a potential mediating factor. Second, we will collect data from pregnant people's EMRs representing three health outcomes (pregnancy-induced hypertension, prenatal depression, and cesarean delivery) that are central sources of pregnancy-related morbidity and have been linked to psychosocial factors and health behaviors (which will also be measured using EMR data).

This aim describes a number of methodological steps, and contained within them is the mention of "open-ended questions" that will collect qualitative data on members of families or features of family life that may have been missed by the network survey approach to measuring family structure and functioning. Note that the balance of the work tips toward the use of quantitative data—surveys and electronic medical records. Also, this aim focuses on data collection, and the second aim of the proposal focuses on analysis. This proposal is for a pilot project that tests the feasibility and payoff of measuring family dynamics through a social network approach, as well as the use of EMRs for health outcome data, so it makes sense that at least one of the aims outlines the collection of primary data to achieve these goals.

In parts of a mixed methods project that involves testing theory, there will also be hypotheses that flow directly from one or more of the research questions guiding the project. Social scientists typically do not explicitly name hypotheses for research involving the analysis of qualitative data.⁴ The qualitative research paradigm leans heavily inductive, meaning the focus is on building theory and hypotheses rather than testing them (Lareau 2012). Also, research involving qualitative data often operates from a grounded theory perspective, in which value is placed on coming to the data without strong expectations for what might result (Charmaz 2014). Therefore, researchers should carefully consider when and how the use of “hypothesis” language is appropriate for strands of research in a mixed methods research project proposal. If your audience (especially reviewers) is not used to or disagree with using hypothesis language for qualitative analyses, it will at best confuse and at worst frustrate readers.

Understanding the important signals sent through research questions (or aims) and hypotheses, such as key concepts guiding the research, relationships between those concepts, and the dependence (or not) of questions/hypotheses on each other, helps mixed methods researchers communicate their objectives, plans for research, and integration points early in a proposal or publication. Having worked out these key components of research questions and hypotheses from the start also provides a clear roadmap for the research itself.

EMERGENT RESEARCH QUESTIONS OR HYPOTHESES IN MIXED METHODS RESEARCH

In most presentations of research findings, written or otherwise, the research questions or hypotheses for all strands of a project are described as if they were crystallized in the early phases of the research. This can obscure how and when the research questions are actually developed or refined in a sequential mixed methods approach. Therefore, it is important to explicitly consider how one strand of research can be used to develop or revise compelling research questions or hypotheses to guide a subsequent strand of research.

In this section, we start by discussing how qualitative data are often used in an exploratory way to develop hypotheses that are then tested with quantitative data. This is followed by considering how deductive quantitative data

⁴ For more on the debate over using “hypothesis” language in qualitative research, see LaRossa (2012) and Lareau (2012).

analyses might produce research questions that could lead to contributions from an abductive analysis of qualitative data. Finally, we discuss the challenges in writing grant proposals for sequential or emergent mixed methods research.

Building Theory to Specify Confirmatory Research Questions

There are moments in the research process when puzzles arise for which existing theory is limited. These puzzles stem from inconsistent or unexpected empirical findings and present opportunities for mixed methods endeavors. Researchers can first conduct an exploratory strand of research designed to produce insights that result in a set of new or revised theoretical propositions. This new theory can then be tested with a systematic confirmatory analysis. This can also happen in an unplanned manner, where the findings from one strand of research spark questions that can be better answered in another strand. The classic approach to this in mixed methods research is to start with the collection and analysis of qualitative data through ethnographic observation and/or interviewing followed by survey data collection and/or analysis.

One example of a mixed methods study in which theory is built through the analysis of qualitative data from intensive fieldwork and then tested using large-scale, representative quantitative data is a study by Mario Luis Small and colleagues on how childcare centers connect parents to important resources (Small 2009b; Small, Jacobs, and Massengill 2008). The motivation for this research was a set of inconclusive findings regarding the relationship between neighborhood poverty and residents' access to economic opportunity. Prior research has focused on individual and neighborhood circumstances to explain this association. At the individual level, social isolation theory has been used to argue that residents of poor neighborhoods have fewer middle-class members of their social networks, cutting down on their access to social and cultural capital (Wilson 1996, 2012). At the neighborhood level, scholars have used de-institutionalization theory to posit that poorer neighborhoods contain less economic, social, and cultural capital in the aggregate. This lack of available capital has been argued to weaken local organizations such as employers or churches, making them less effective at helping residents (Wilson 2012). Tests of these theories have produced inconsistent results (Goering and Feins 2003).

Small, Jacobs, and Massengill (2008) point out a lack of attention, in this well-developed literature on the effects of neighborhood poverty, to networks of local organizations that may serve residents by connecting them to the

resources they need. Thus, they launched a study with an “open systems” theoretical orientation and an abductive, ethnographic empirical approach designed to reveal specific mechanisms through which childcare centers and their organizational ties offer resources to patrons. The research questions for this first strand of their research were, “Do childcare centers provide access to resources important to well-being through their inter-organizational ties? If so, how is this process affected by neighborhood poverty?” (Small, Jacobs, and Massengill 2008:390). They knew that answering these questions would produce the insights needed to specify new research questions answerable in a second, confirmatory strand of research.

The first strand of this mixed methods project involved in-depth observations and interviews with childcare center directors, other personnel, and parents who were clients at 23 childcare centers in four New York City neighborhoods selected to achieve diversity in racial-ethnic composition and income levels (Small 2009b). Key informants who were government or non-profit sector leaders were also interviewed.

Analysis of the resulting qualitative data suggested that indeed childcare centers, through multiple actors, provide a wide range of important resources to clients through their ties to other organizations. The intensive fieldwork focused on the voices and lived experience of childcare center staff, parents, and key informants in linked organizations. The analysis was abductive because it focused on a hunch that an organization’s relationship to other organizations might connect patrons to further resources. However, it also allowed those relational processes to be revealed as they were rather than starting with a well-developed theoretical model.

Small, Jacobs, and Massengill (2008) discovered multiple forms of informational, service, and material resources that childcare centers provide parents. They found that sometimes childcare centers referred patrons to other organizations and sometimes the centers collaborated with other organizations to jointly provide resources. The fieldwork also revealed key contingencies shaping the organizational networks of childcare centers. They found that government-funded centers were mandated to provide multiple services to families, so they were obligated to develop and activate ties with other organizations. Also, large nonprofits were often motivated to fill gaps left by government services, so they worked to link childcare centers to resources not provided by the government. Thus, the type of childcare center and its collaboration with large nonprofits altered the types of interorganizational ties they formed. The research team also found that the degree to which center directors embraced professional norms around a “holistic” approach to childcare influenced the emphasis they put on tie formation and activation. In addition, the perceived poverty of

a neighborhood raised the likelihood that public agencies or private nonprofits would provide resources to its childcare centers.

The many insights produced by the first strand of the study facilitated the development of specific hypotheses about how neighborhood organizations link community members to resources and whether those processes vary by neighborhood socioeconomic level (see Figure 2.1). A second strand of research was designed to specifically test these hypotheses with survey data from a representative sample of childcare centers in New York City ($n = 293$). The survey questionnaire included measures of many of the key concepts and processes discovered during the earlier fieldwork. Taking a more objective, deductive, and statistically generalizable approach, this second strand of the mixed methods project allowed the research team to systematically confirm and extend many of the findings from the first strand of research.

Using an exploratory to explanatory sequential approach to mixing methods does not always mean combining ethnography and survey methods, although this style of mixing is common. Rather than starting with ethnography, one might use semistructured interviews on their own either one-on-one or with a focus group approach to construct and/or refine a theory that is

FIGURE 2.1 ■ Hypotheses Developed Through Ethnography for Testing With Survey Data

Hypothesis 1

- Because of their organizational ties, childcare centers will provide access to multiple resources important to well-being.

Hypothesis 2

- For-profit centers will exhibit the fewest and government-funded centers the greatest number of active resource-rich organizational ties.

Hypothesis 3

- Despite H2, sector differences will not fully account for the variance in number of ties — even after controlling for the center's sector, centers serving poor patrons will exhibit a greater number of ties than centers not serving the poor.

Hypothesis 4

- After controlling for the poverty level of the patrons served, centers in high-poverty neighborhoods will still exhibit more ties than those in other neighborhoods.

Note: Adapted from “Why Organizational Ties Matter for Neighborhood Effects: Resource Access through Childcare Centers,” by M. L. Small, E. M. Jacobs, and R. P. Massengill, 2008, *Social Forces* 87(1):387–414 (doi: 10.1353/sof.0.0079).

then tested through statistical analysis of survey data, existing administrative data, or results from an experiment. Regardless of the type of data collection, there is much precedent for using the ideas arising from abductively analyzed qualitative data to craft and specify research questions or hypotheses to be assessed with some form of quantitative data.

In order to implement this approach, researchers should develop clear theoretical models from their analysis of qualitative data that can be tested with survey data. This means that the concepts must be ones that can be operationalized with survey questions and must be defined clearly in order to implement this approach. In the research from Small and colleagues, for example, the concepts derived from qualitative work included type of childcare center, ties to other organizations, resources given to parents, family poverty, and neighborhood poverty. All of these concepts can be measured directly by survey questions. On the other hand, researchers must avoid either attempting to fit the collection of quantitative data to an ill-defined theory or using survey questions that are not precise measures of the concepts derived from the qualitative data.

Revealing the General Context to Specify Research Questions

Flipping the types of data and analysis around, another sequential study design is to use the analysis of quantitative data to reveal patterns or associations for further examination with the analysis of some form of qualitative data. The quantitative data analysis might produce expected or surprising findings that spark new research questions or hypotheses. One exemplar of this approach is a study by Howard and Roessler (2006) about the process through which an incumbent authoritarian leader or party comes to lose an election. They start by using a set of well-justified criteria to identify 50 competitive authoritarian national elections held between 1990 and 2002. They create a data set including key characteristics of the context in which these competitive authoritarian elections were held, and they run logistic regression analyses to see which characteristics are associated with the odds of an authoritarian incumbent leader or party losing the election, or what they label a *liberalizing electoral outcome* (LEO). They find that structural forces such as recent economic growth, foreign direct investment, foreign aid, parliamentarism, and regime openness are not as important factors as the strategies used by the opposition. When an opposition candidate or party develops a coalition with other parties and mobilizes, an LEO is more likely.

Howard and Roessler's (2006) findings about opposition party strategies then led them to pose a new research question: *How* do opposition party

strategies come to matter? This kind of question, which asks how something happens, is often best answered by analyzing qualitative data. In this case, Howard and Roessler take a “nested analysis” approach (Lieberman 2005:435). They select the case of Kenya’s 2002 election of Mwai Kibaki over Uhuru Kenyatta, who was the appointed successor of Daniel Moi. Using this case, they perform an in-depth analysis to trace the processes resulting in this particular LEO. Using qualitative data from records and reports, they provide a detailed accounting of how the broad-based National Rainbow Coalition (NARC) was formed and mobilized despite longstanding personal and ethnic differences between members. They argue that it was not the strength or the popularity of the opposition party, because they received around the same proportion of the presidential vote as they had in prior years. Instead, it was their methods of organization. By bringing together so many opposition parties in a coalition, they were able to channel the opposition vote into one candidate. They also raised the risk and cost of the use of fraud by the incumbent party by developing an early lead in the polls. Knowing a Kibaki victory was likely, Kenyatta’s party, which was already running on a campaign to end corruption and could not risk being vulnerable to legal sanctions if they lost, did not resort to prior forms of manipulation and rigging. This made the 2002 election one of the fairest in Kenya’s history, paving the way for a Kibaki victory.

In this example, we see that the analysis of quantitative data about a wide variety of elections was necessary to establish the strongest correlates of the likelihood of an LEO. Once these key correlates had been empirically confirmed, the researchers could more certainly form the question guiding the nested, in-depth case analysis and process tracing. Specifically, the authors here needed empirical evidence for the role of oppositional strategies to then formulate questions about *how* those strategies achieved success. If they had instead found that structural factors, such as foreign aid, were important, they would have focused the question for the second strand of research, the nested case study, on explicating the mechanisms for those forces. Thus, this study is a nice example of a sequential analysis in which the quantitative data and analysis allow researchers to then formulate a specific research question for follow-up and extension of the initial findings.

Funding Emergent Strands of Sequential Mixed Methods Research

When conducting a sequential mixed methods project where the design of a second- or higher-order strand depends on the findings from a prior strand of research, it is not automatically clear how to propose funding for the entire project at once. Good proposals have a clearly articulated set of research

questions or aims. Reviewers would not be satisfied with a proposal that lays out the details for one strand of research and then says the details of the second strand will be worked out later (based on the findings from a prior strand of the project). There are two options for this kind of mixed methods study. First, you could seek funding separately for the two strands. This allows you to wait until you are more certain of the research questions for the second strand to request funding for that part. Of course, this will take more time because review and revision cycles can be lengthy. The second option is to anticipate a likely set of research questions, hypotheses, or aims that might come from the analysis of data in the first strand of research and build a plan for the second strand of research off of that. You should explicitly state that you will remain open and flexible to whatever the outcomes of the first strand of research are, and that may require revising part of the plan for the second strand, but if you can produce a well-designed hypothetical second strand, reviewers will likely trust you to make those adjustments in the field. In other words, a well-reasoned proposal, combined with a stated commitment to being flexible in the field, will go a lot further in giving reviewers confidence in your project than a vaguely worded promise that you will do your best to develop research questions/hypotheses/aims for strand two when you get there.

APPLICATION EXERCISES

1. To develop a research question (mixed methods or not), it is helpful to work through several steps:
 - a. Identify a topic in which you are interested.
 - b. Within that broad topic, select a specific issue and population in which you are interested.
 - c. Brainstorm a list of research questions about that narrower topic.
 - d. Eliminate questions that you are less interested in or are not feasible.
 - e. Take the three most interesting questions and revise them to make them specific enough to answer.
2. For each research question, what kind of quantitative and/or qualitative data could you collect to answer it? If both kinds of data would work, reword the question so it would work as a single, overarching research question. If one of your questions could best be answered by only one type of data, come up with some alternative follow-up questions that could arise, depending on the findings from your first strand of research.

Do not copy, post, or distribute