

Data Quality in Multi-sited Cross-Sectional and Panel Studies.

By: Eric C. Jones, Arthur D. Murphy, Julia L. Perilla, Isabel Perez-Vargas and Fran H. Norris

[E.C. Jones](#), [A.D. Murphy](#), J. Perilla, I. Pérez-Vargas, F.H. Norris. Data Quality in Multi-sited Cross-Sectional and Panel Studies. *Field Methods* 22(3) (2010):250-269.

Made available courtesy of SAGE Publications: <http://fm.sagepub.com/content/22/3/250>

*****Reprinted with permission. No further reproduction is authorized without written permission from SAGE Publication. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document. *****

Abstract:

The authors address the issues faced while collecting survey data as part of a large multisite, multidisciplinary long-term project using interviewers rather than self-administered questionnaires in a country in which the researchers are not native. The issues pertain to the collection of high-quality data that accurately measure the variables of interest from which generalizations can be made. Three issues were prominent: potential cross-cultural variation in the validity of measures; how to manage multiple control sites and multiple study sites; and how to control for problems presented by series/panel data (i.e., the influence of prior interviews or subsequent intervening events on later interviews). The authors addressed these issues through five strategies at different points in the study. This discussion concerns the challenges and benefits of using these techniques to address the three main issues.

Keywords: anthropology | psychology | disaster | Mexico | data quality | data collection | research methods

Article:

Introduction

In October 1999, a stationary tropical depression in the Gulf of Campeche generated torrential rains, widespread flooding, and devastating mudslides in nine Mexican states, killing more than 400 people and forcing at least 200,000 people from their homes. Before this, our team had begun a series of studies, including a cross-sectional epidemiological study of four Mexican cities to provide a normative sample for mental health. That study would eventually be followed by a separate longitudinal panel study of a Mexican disaster, which turned out to be the 1999 flooding of the Gulf Coast of Mexico. This article discusses both studies. To understand how the 1999 flooding event was experienced, we conducted a four-wave panel study of two Mexican communities: Villahermosa, the capital of the coastal state of Tabasco, population 500,000, and Tezuitla'n, a mountain city in the state of Puebla, population 60,000. The severity of this event made it well suited for studying the psychological consequences of disasters in Mexico and for

testing the cross-cultural applicability of a theoretical model (the Social Support Deterioration Deterrence Model; see Norris and Kaniasty 1996) that guided our previous disaster research.

In this article, we address the issues we faced while collecting data as part of a complex multisite, multidisciplinary panel study in Mexico using interviewer-administered questionnaires. Specifically, three issues were prominent in the collection of high-quality data: potential crosscultural variation in the validity of measures; how to manage multiple study sites; and how to address problems presented by the influence of prior interviews on later interviews in longitudinal research. In our study, we addressed these three issues using five strategies: (1) designing the study with primary attention to the order of data collection events, (2) systematic and intensive training, (3) using culturally appropriate consent and compensation, (4) checking data in the field, and (5) analyzing data for completeness.

Issues in Managing Complex Projects

Several issues make acquiring quality data challenging when doing survey research. We have found most can be addressed with the strategies we have enumerated. For example, a large sample size can result in challenges because of design sampling errors, noncoverage of a population, nonresponse, and inaccurate response (Groves 1987), the last of which can be a result of interviewer effect (see also Assael and Keon 1982). Feskens et al. (2006) suggest improving contact rate and making questionnaires more understandable to reduce nonresponse and inaccurate response, and McCarty and Killworth (2007) emphasized how significant the role of burden on the informant can be in determining the response rate, although greater contact effort sometimes pays off. Laurie et al. (1999) suggested six contacts before replacement or dropping. In addition, bigger studies require more people to collect data, which creates greater possible interviewer effects as well as variation in quality of training and collection procedures. Interviewer effect can also occur as a result of the layout or graphic format of the questionnaire, regardless of interviewer experience (Sanchez 1992). To deal with these sampling and nonsampling errors/biases, training of interviewers (#2) is of considerable importance, as are pretesting questionnaires (#1), checking data in the field (#4), and analyzing data for completeness and errors/bias before using data to test hypotheses (#5).

In addition to size of study, the addition of the longitudinal disaster panel study to the cross-sectional epidemiological study made our study challenging. Regarding multisite studies, cultural and linguistic differences between sites may exist, which can affect generalizability/external validity (e.g., Pasick et al. 2001; Neumayer 2002). Conducting research outside the investigator's home country or with multiple groups of people with potentially different cultural backgrounds presents challenges because of cultural relevance of the standardized questions, cultural differences in concern for issues concerning human subjects, and various expectations regarding reciprocity, or what the researcher brings to the table. Research not involving door-to-door interviewer-mediated surveys has shown that incentives or desire for incentives can affect response rates because of level of individualism, perceived burden, level of exposure to surveys

in general, and political powerlessness (Storms and Loosveldt 2004; Heerwegh 2006; van den Brakel et al. 2006). Our surveys were interviewer administered, and the review of Singer et al. (1999) indicates that high initial response rates without incentives—which was the case for our study—means incentives will produce small increases in response rate and answer completeness. To address the above issues, culturally appropriate consent and compensation (#3) are necessary, as are pretesting (#1) and appropriate training (#2).

Panel studies are relatively rare in anthropology, and a recent review paper calls on anthropologists to make greater use of this methodology especially because of the ability of panel data to address causal relations (Gravlee et al. 2009). Panel studies are also subject to several data quality issues. Regarding the effect in panel studies of time 1 interviews on later interviews (resulting in potential bias), a review by van der Zouwen and van Tilburg (2001) stated that memory effects of the interviewee are unlikely and conditioning effects or consciousness raising may occur in studies of political participation, but this tends not to happen in other types of studies. However, panel studies can cause the participant to alter responses because of interviewer effect. Interviewer effects involve developing particular expectations about typical responses, typical survey duration, and probable response based on prior interviews with the same protocol and/or same person (van der Zouwen and van Tilburg 2001). Van der Zouwen and van Tilburg thus suggest hiring new interviewers for each wave of questionnaires as well as not providing interviewers with data from prior interviewers unless methodologically necessary. They also suggest: strict training of interviewers (#2), maintaining reasonable work loads so interviewers are not rushing (#2), and closely monitoring interviewers (#4), even using audio or video recording to analyze interviewer actions.

We used the above techniques, as discussed throughout this article. In relation to our particular study, we first take up study design, then training, consent and compensation, data checking, and finally quality aspects of data analysis before discussing further generalizable challenges of multisite studies. This methodological investigation does not perform tests on the quality of the data collected but covers pluses and minuses of using various techniques to collect the best possible data for multisite and panel studies.

Study Design

For our study on the effects of trauma on mental and physical health in Mexico, we put together a team of community psychologists, clinical psychologists, sociologists, demographers, and anthropologists from both the United States and Mexico and believe the multiple perspectives of the team members helped us avoid potential problems with data quality. This was certainly the case in the preparation of the data collection procedures. Because we sampled multiple sites over the course of several years, we first established the domain of our study through preparatory qualitative and then quantitative data collection. We then pretested our instruments before conducting the cross-sectional epidemiological surveys and the four-wave panel disaster surveys.

Table 1 is omitted from this formatted document.

Preparatory Studies

Before embarking on the larger cross-sectional epidemiological survey and the panel disaster survey, we conducted qualitative and comparative research to better understand the domains with which we would be working in Mexico. The first project was a qualitative study of posttraumatic stress among Mexican disaster victims in one site in the United States and two sites in Mexico in which the interviewees described symptoms and we assessed the extent to which those symptoms overlapped with Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) criteria symptoms (Norris, Weisshaar et al. 2001; see Table 1). Next, we conducted a cross-cultural study on disaster-related posttraumatic stress disorder (PTSD) in Mexico and the United States resulting from hurricanes Paulina and Andrew, respectively (Norris, Perilla et al. 2001). The results indicated that post disaster stress is a meaningful construct in Mexico and we came to understand some cultural variation in the different PTSD criteria. Results from these preliminary studies prodded us to further investigate disaster mental health in Mexico, using both a panel study of disaster-affected cities (e.g., Norris et al. 2005) and a cross-sectional epidemiological study (e.g., Norris et al. 2003) of four unaffected cities to serve as a normative comparison.

Pretest

The interview protocol was composed of a broad array of measures of psychological, physical, social, ecological, and economic well-being. We used the Composite International Diagnostic Interview (CIDI) Version 2.1 (World Health Organization 1997) to provide DSM-IV assessments of lifetime PTSD, major depressive disorder, generalized anxiety disorder, and panic disorder. The English version of the CIDI has been used widely in prior epidemiological studies, although we know of no validation that has been performed on the World Health Organization's Spanish version.

We pretested the instrument in Oaxaca City, Mexico, before starting the training process and administering the questionnaire. Individuals involved in the project represented the regions of Mexico that were to be studied, giving us some assurance that local variation would not cause major problems.

Epidemiological Study

A multistage probability sampling design was used to draw representative samples of adults from the four cities. Using Mexican census data, we randomly selected 10% of the census tracts in the urban core of each city. For example, in Oaxaca, we selected 24 census tracts and, from those, 903 households of which 727 were eligible for the study. Between zero and three houses were interviewed in each block, as some blocks were not eligible (businesses or vacant lots) and larger blocks had proportionately more interviews. In each block, the interviewer started at one of the

following corners: northeast, northwest, southeast, and southwest by rotation with each subsequent block. The interview was conducted with the third house to the right of that corner. If the sample required that a block have more than one house to interview, then the second interview was conducted with the third house to the right of the first. If a household refused to participate or was empty after three tries, then the interviewer went two houses to the left to conduct the interview and two more to the left if that household refused, continuing to the left. If no one was at home at the selected houses after three tries for all of the houses in the block or each selected house refused, the block was replaced with a block randomly selected again.

To reduce serious ethnic confounders into the study, we drew our sample from the urban core, as the study was designed to focus on Mexican mestizos (people of mixed European, indigenous, and occasionally African blood) and not on indigenous populations. We worked with local expert informants to determine what sections of the cities in the study had not experienced massive recent immigration from indigenous regions. At the initial household contact, the adult who answered the door was asked to participate in a sociodemographic interview about the household. Then, among household residents, one adult was selected using a randomization process based on the person who had most recently had a birthday. The one adult in each household was then asked to participate in the study by being interviewed at a later date. Sample size and response rates for each city in the cross-sectional epidemiological study were 576 (79%) in Oaxaca, 713 (82%) in Guadalajara, 618 (76%) in Hermosillo, and 602 (70%) in Mérida, for an N of 2,509. The Oaxaca and Guadalajara data were collected in 1999, the Hermosillo and Mérida data in 2001. It was necessary to replace some people in the sample in each city to achieve adequate ns, although we did not count the replaced households as responses when we calculated those response rates above. The male–female ratio of 45:55 was representative of the Mexican census results.

Interviews were completed by trained, local interviewers. In each city, we recruited interviewers from local universities. We did not limit candidates to a particular discipline; the general and most important requirements were a willingness to learn and a willingness to talk with people. The lengthy training process gave us opportunity to note how people would perform with the interviews and related challenges. Interviewers were paid a sufficient amount per completed interview so that an individual could earn approximately twice the average wage in a normal work day. No one was paid for an interview until it had been checked for internal consistency and any unanswered questions or inconsistencies verified in the field. Fieldwork managers checked all interviews for accuracy of selection procedures, completeness, and quality.

Panel Study

Initial interviews were conducted 6 months postdisaster, in April 2000. From affected census tracts in Villahermosa, 653 households were sampled randomly in proportion to the tracts' population sizes. Of these, 461 completed the psychological interview, for a final wave 1 response rate of 77% of eligible units and 87% of those actually contacted. In Tezuitla'n, all 235

households provided with plots in the relocation community were selected and, of these, 205 participants completed the psychological interview, for a final response rate of 87% of those eligible and 98% of those actually contacted. Attempts were made to reinterview all participants at 12, 18, and 24 months postdisaster. In Villahermosa, 385 or 84% of the wave 1 participants completed all four psychological interviews, as did 176 or 86% of the wave 1 participants in Tezuitla'n, for a final four-wave N of 561.

All interviews were completed in respondents' homes, with care taken to ensure confidentiality of responses. We used interviewers with limited or no prior experience with the survey instrument, thus reducing the possibility that, based on previous experience, there would be an a priori expectation as to how long it would take to complete the instrument. This was further reinforced by conducting several practice sessions, some of which included complicated cases requiring the maximum estimated time to complete.

Training

Because interviewer effect is the source of most data quality issues in panel studies, training was a high priority in this study. Good fieldwork managers are crucial and it is usually best if they are locals. Training is crucial in all research, not just in very complex studies. We examine our training because of the particular challenges produced by conducting trainings in numerous and varied settings.

First, all the interviewers were brought together to meet one another and to hear a detailed explanation of the study and its objectives.

Second, we explained every single question, why it was being asked, and how to ask or read each one. Here, we found it necessary to explain the importance of following the protocol exactly and providing proper delivery, plus the possible errors that can result if this is not done.

Third, we paired interviewers together to administer the questionnaire to one another and then discussed the issues they brought up and questions they had. Again, it is imperative to emphasize the importance of the interviewers having an understanding of, and rationale for, every question they were asking.

Fourth, we explained the sampling procedure to the interviewers. This turned out to be one of the most difficult parts of the training. Although the sampling procedure described earlier might seem straightforward, we found there was a difference between hearing how to sample versus being able to explain how to sample versus actually getting out and consistently doing it correctly.

And fifth, when we had a majority of people who understood the questionnaire and/or sampling procedure, we asked them to go to the field and start conducting the survey. After practicing for a couple of days, they came back for further discussion and training.

We started out with many more people collecting data than we needed for the study. We knew some interviewers would decide the job was not right for them. People left because the work was not easy and because they were not doing an adequate job of conducting the survey. It was important to us to minimize interviewer effect, which required intensive and time consuming training as well as a highly standardized questionnaire.

A final note on training: Our research was about mental health and the potential causes and mitigating factors of mental health problems. Even in the control cities, there were reports of trauma and abuse, and this began to take its toll on the interviewers. The research coordinators took caring and adequate measures to address problems of this sort that arose among the interviewers under their supervision, and we scheduled check-ins or counseling sessions to help the interviewers deal with the stress of listening to so many stories of trauma and/or need. These support sessions occurred about 1 month to 6 weeks into the research in each city.

Culturally Appropriate Consent and Compensation

Other issues that could possibly affect data quality include institutional requirements (in addition to creating an Institutional Review Board [IRB] in Mexico from scratch) regarding privacy, informed consent, and the traditional manner of compensation in mental health studies in the United States.

First, the methodology required privacy, but this was not always easy to achieve. Although the intent of the requirement of privacy is that people can speak freely and truthfully, in Mexico, if people are isolated from family for the interview, they may feel like they are hiding something just by talking to us and thus may choose to speak less or even lie. The important factor for data quality here, then, is not privacy per se, but how the interview setting is handled so people feel comfortable in a private interview.

Second, informed consent was required by the funding agency, and by our host institutions, although early on our collaborators saw it as unnecessary. In Mexico, participation is seen as informed consent—people are not used to many contracts in their lives, and they tend to distrust signature collection. As a result, informed consent required creativity so as not to create distrust of the interviewer and thus potentially affect data quality. We used oral consent, which seemed to put people at ease, while still completing our responsibilities.

Finally, individual compensation is a normal part of much of social science research, more so in psychology than in anthropology. Our belief was that compensation should be a way to say “thank you” rather than an incentive, thus we promised a certificate of participation before the interview and also placed money (1 day’s wage) into the envelope with the certificate so the participant did not know they would receive money until they had completed the interview. Our high response rates in all six cities suggest that the burden was not perceived as unreasonable and that monetary incentives were not necessary or may not have had an appreciable effect on response. Similarly, the communalistic orientation of Mexican society compared to United States

and European research sites probably makes individualistic orientation and political powerlessness less important for needing incentives. We continue to be skeptical about the use of money incentives in this kind of fieldwork, not only for data quality effects but also for the social dynamic of jealousy, perceived favoritism, and sense of injustice that can develop from not including everyone. In such cases, we have found that a donation of equipment to a community, such as the local school or municipal office, serves as a recognized “thank you” from the project.

Data Checking in the Field

Checking Questionnaires

The site coordinator in each city checked each questionnaire as it came in from the field, in both the cross-sectional epidemiological survey and the post disaster panel survey. If any data were missing, the interviewer was asked to go back and complete the interview. We also checked for internal consistency between interview questions. If the answers did not match, we returned to ask for clarification.

When permission was given, all interviews were tape recorded. Early in the interview process, supervisors listened to all tapes as they reviewed the instrument. This allowed us to check both the performance of the interviewer in the field and to check for any systematic errors being made by any individual or the group as a whole. Once satisfied that there was no systematic problems with an interviewer’s questionnaires, supervisors spot-checked interviews against the tape. If there were any internal inconsistencies in the instrument, the tape recording was used to attempt to resolve them before sending the interviewer back to the field to check with the participant. Typically, what appeared as an inconsistency was resolved by the interviewer who would have received some qualitative information in the course of the interview. If not, the interviewer was asked to return and clarify any apparent inconsistencies.

The tape recordings were useful for checking for panel conditioning, memory effects for the interviewee, and other anomalies. In addition, our analyses of the panel data concerning current levels of depression and trauma for each of the four survey waves showed a general decrease over time in symptomology, suggesting to us that people were answering about their current state.

Finally, we took a 10% random sample of all completed questionnaires and returned to see whether the interviewer had talked with the correct individual. Within 2 weeks, interviewers were turning in good work or had quit.

Each questionnaire had a control sheet on which we recorded information such as interviewer name, date of survey, time begun and finished, where the questionnaire had been administered, if and why a house or an individual was replaced, and the times and days interviews were attempted (up to three times).

Checking Data Entry

For data entry, we hired someone who had entered data for the Mexican Census Office. After data entry, we produced descriptive analyses of the data to see what the range of variation looked like and to see if there were any outliers that could be explained by typing errors. Very few errors occurred, but we were able to catch some. We discussed having someone else check every 10th or 20th survey against the entered data but decided to forgo this strategy, given the high accuracy of the data entry process.

Because epidemiological studies in the United States commonly rely on computers for data collection, we experimented in a small pretest with computer-assisted interviews to eliminate data entry and potential error in data entry. However, we had to abandon the approach because of the local interviewers' discomfort with conducting the surveys this way. We were also limited by inadequate battery life for the laptops we were using.

Analysis of Data Quality

Good training and data checking in the field should be accompanied by analysis of data quality to assure reliability of results. We can ascertain whether our individual samples in each city approximated the household samples from which they came, on such variables as gender, age, education level, and civil status. Variation in dependent variables between cities may exist and may be due to a number of things. However, we leave for another article the analysis of such variation. Instead, we discuss the likely culprits for sampling error (i.e., dissimilar variation in samples).

Sampling error (68% females instead of 55% as per census) might have occurred as a result of using households as the preliminary point of selection instead of a list of all adults. Sampling error could also have occurred through the randomized selection of the blocks to be surveyed, although the effect of the sex of respondents on that aspect of sampling is not intuitive. Aside from sex as part of such a bias, our choice of what blocks were considered urban may inadvertently have included more suburban-type populations than desired or we even may have oversampled the less dense blocks in the city. Noncoverage of a population is unlikely because of our sampling procedure, although some noncoverage was intentional—we purposefully excluded suburban dwellers, indigenous people, extremely wealthy individuals, people without homes, and children under 18 years of age. Nonresponse indeed occurred in some of the later questions in the long survey. Inaccurate responses were mitigated using scales with 4–30 questions rather than individual questions as independent variables.

Two other minor points are worth noting concerning quality of the sample. First, we think the low level of missing data in our study—and thus less potential skewness in data—occurred not only because of the appropriate training of the interviewers but also because of the general willingness on the part of the participants to participate in and complete the interviews. Second, for the follow-up visits to 10% of the interviewed households to see if data were collected

correctly, we unfortunately did not record the results of follow-up visits—so we do not know if there were systematic errors/bias, nor do we know what the rates of error were. It is our sense that follow-ups were effective as a part of the data quality control process.

Discussion

Through the above techniques and strategies in design, training, cultural interfacing, data checking, and data analysis, we sought to deal effectively with three major issues typically experienced in collecting quality data in multisite studies. These issues are (1) cross-cultural variation between study sites resulting in differential interpretation of the meanings of questions asked, (2) choosing sites that vary appropriately to improve generalizability, and (3) dealing with effects of longitudinal issues. The specific concerns about representativeness or sample quality (along with validity of measures) come from concerns about generalizability, a fundamental goal in most sciences. In addition to the obvious potential problem of a sample not being representative of the population, when does variation in the distribution of gender, age, and civil status across sites or populations affect generalizability? And when do differences between self-reports versus other sources of measurement of dependent variables (e.g., clinical measures) have an impact on the quality and generalizability of the results of the analysis?

Cross-Cultural Variation in Validity of Measures

Appropriate local language for questions has the biggest influence on generalizability or external validity in cross-cultural research (e.g., Pasick et al. 2001). As mentioned previously, our questions were not tailored to local language norms, nor were interviewers expected to follow up with extensive explanatory prompts—although anecdotally we know that interviewers at times prompted participants to avoid missing data. We think having locals helps, so that the nonscale questions—even though we had fluent Spanish-speaking team members from the United States—come across as intended. Although we pretested the protocol, limited the sample to the core areas of major urban areas in a single country, and used local interviewers, it is possible that local language differences exist that affected our study.

A few other factors can also come into play that might limit the extension of findings from a single country or single culture study. First, how willing are people to respond to potentially taboo questions or to give shameful answers? Even when people respond to questions, some subjects are just not discussed in a straightforward truthful manner, so answers may be inaccurate. In addition, the intonation of the interviewer might cause a person to skew their answer, although we sense that intonation was perhaps less important than interviewer body language—making the participant feel comfortable or not. In the end, it seems that data quality control can depend on the researchers' and the interviewers' knowledge of the study population.

One other problem that might be addressed through knowledge of local society concerns sampling, as geographic distribution of people in the study site might not fit sampling

assumptions. Sampling techniques would thus be improved by an understanding of local residence pattern and household systems.

How did interviewers know if they were given accurate data? Sometimes interviewers wondered whether they were given good data or not and, although interviewers were often initially naive about the research process, they generally asked insightful questions of us and became comfortable with material through training and repetition. They became researchers invested in the data and results and they felt that people were willing to be open and answer their questions.

Choice of Multiple Study Sites

The use of multiple study sites usually increases the generalizability of data. However, we assume that the choice of those sites is made on the basis of a theoretical framework—in our case, capturing the range of variation across large cities in Mexico with the objective of understanding the mental health of economically diverse urban dwellers (cross-sectional epidemiological study) and how their mental health is affected by disaster (panel study). In consultation with our Mexican colleagues, we chose geographically dispersed sites representative of relatively large Mexican cities with varying trajectories of economic development, then our sampling was devised to avoid suburban areas. Random sampling of comparative sites provides the greatest generalizability, but the cost of such a design may be prohibitive, and sufficient variation in the study variables might not be achieved without a stratified sample. However, we knew this was an once-in-a-lifetime opportunity to conduct a study that would have an accurate urban normative sample for an entire nation and that would also test variation in post disaster mental and physical health in urban areas compared to epidemiologic norms.

Longitudinal Studies and Internal Validity

The internal validity of a questionnaire in a longitudinal panel study involving comparison between geographically distinct populations faces the same challenges as does noncomparative research, including longitudinal changes in interviewing factors or in a dependent variable not due to changes in an independent variable; contaminated responses based in increased familiarity with the questionnaire when administered multiple times (on the part of the interviewee or the interviewer); changes in how survey questions are presented; and attrition of interviewees in subsequent interviews.

These longitudinal issues should not pose greater problems for comparative projects than for single-site projects. For longitudinal changes in independent variables, an ethnographic understanding of the study site helps with the formulation of alternative hypotheses—rather than assume that changes in a dependent variable occurred because of changes in an independent variable. Thus, the project managers also conducted ethnographic work in the two cities in which we collected longitudinal data. Project managers lived in each of the two panel study cities for the duration of the studies (just over 2 years), as did the local interviewers, although the latter

were not involved with the study except during the 2 months every 6 months when surveys were conducted.

The type of survey we administered is typically not subject to contamination from prior exposure to the instrument because it is very long, relatively complicated to administer and respond to, and does not contain political material (e.g., van der Zouwen and van Tilburg 2001). Thus, informant memory of answers is expected to be low, and attempts by the interviewee to second guess the objectives of the question would be quite exhausting to the interviewee. Some interviewer effect may result because of expectations regarding the answers to some questions or the time required for completing the interview, thus the suggestion of van der Zouwen and van Tilburg (2001) to hire new surveyors for each wave. However, Laurie et al. (1999) advise using the same interviewers to return to same interviewees, if possible, to maintain high response rate and low attrition. We opted for the latter—a higher response rate through using the same interviewers—and tried to limit interviewer effect by not providing prior interviews to the interviewers. The importance of the first interviewer in a panel study for subsequent participation cannot be overstated (Pickery et al. 2001), thus underlining our efforts in the training process to make that first contact effective, including carefully explaining the process of participation. In addition, interviews occurred 6 months apart, so interviewer recall should have been relatively minimal concerning most answers of any individual respondent.

Another solution to interviewer effect, panel conditioning, and respondent burden/attrition was proposed by Munger and Loyd (1988; see also Adams and Darwin 1982; Zeger and Thomas 1997). The procedure is like this: First, administer section 1 (basic demographic info) to everyone, then sections 2, 3, 4, 5, and 6 to the first 100 people, then sections 3, 4, 5, 6, and 2 to the next 100 people, and so on. That helps control for possible participant fatigue or dropout, although it can increase the N necessary to assure that the margin of error for each variable is acceptable. Now that there are good tools available for imputing missing data (e.g., Raghunathan and Grizzle 1995; King et al. 2001), such a design would likely have improved the quality of our data. Although our questionnaire had several clear sections, we did not use this design to reduce fatigue because of our fear that interviewers would get confused when having to administer several versions of the survey in the same day.

Changes in how questions are presented because of differing personnel are worthy of consideration because of the large effects it may have on data quality. We were not able to test for inter-interviewer variability in our data because we did not input those data into the computer before we destroyed the paper surveys, as per our IRB proposal. We had multiple interviewers in each city—but no interviewers worked in all the cities—and we conducted our analyses across all cities, so any effect that did occur would not be systematic across cities. Because we used the same interviewers in each city, we did not have random measurement introduced by changing personnel and their wording of question. In addition, when reviewing tapes, we checked to see that questions were being asked correctly.

Finally, interviewee attrition might also cause unwelcome variation in data and data quality. We believe high retention of participants in our study avoided that problem. We attribute high retention of participants partly to good training and the skill of fieldworkers (professionalism, courteousness, listening well, efficient use of time, etc.) but also to factors external to the study, such as the cultural openness of both the study and control populations as well as relatively low transience of the study populations. The average age of study dropouts was 37 compared to 35.4 years of age for people who stayed in for all four waves. Average years of education for also higher (8.4) for those who quit versus 7.9 years for those who remained participants. These differences are significant but explain very little of the variation. Negative affect was not statistically different for those who left the study versus those who did not leave (7.7 vs. 7.9), and gender was also not a significant predictor. Thus, the limited attrition we did experience appears to be of a relatively random sort. We did weight the sample for gender because of sample bias but not because of any other factors. Vandecasteele and Debels (2007) showed that weighting based on attrition must use longitudinal weights that include burden, nonresponse, interview experiences of interviewees, and education level as predictors of dropout.

Conclusion

Data that are representative and that can tell us something about change are expensive and hard to produce. However, regardless of study scope, effective techniques can be used at various points in a study—protocol design, training, consent and compensation, and checking data in the field, analyzing data for representativeness—to maximize quality of data. A more specific list of those techniques would include effective training, going back to check a subsample of interviews, controlling for interviewer effect, checking data entry, considering cross-cultural differences in measure validity and effects of sampling strategy, changing the order of questions from interview to interview, perhaps hiring new fieldworkers in each stage of longitudinal research (but keeping the same project managers and presentation of question prompts the same), minimizing informant attrition through appropriate consent and compensation and, in analysis, controlling for changing contexts or changes in dependent variables not due to changes in independent variables used (see Laurie et al. 1999 for additional techniques for maintaining high longitudinal response rates).

Acknowledgment

We would like to thank Mirta Isabel Figueroa, Fabiola de Anda Pe´rez, and Julieta Tello Bello for their dedication as fieldwork managers throughout the project.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: The research upon which this methodological investigation was based was supported by Grant No. 2 RO1 MH51278 from the National Institute of Mental Health, Fran H. Norris, principal investigator, Arthur D. Murphy, coprincipal investigator. The College of Arts and Sciences at the University of North Carolina at Greensboro provided additional financial support for the write-up of this research.

References

Adams, L. M., and G. Darwin. 1982. Solving the quandary between questionnaire length and response rate in educational research. *Research in Higher Education* 17:231–40.

Assael, H., and J. Keon. 1982. Nonsampling vs. sampling errors in survey research. *Journal of Marketing* 46:114–23.

Feskens, R., J. Hox, G. Lensvelt-Mulders, and H. Schmeets. 2006. Collecting data among ethnic minorities in an international perspective. *Field Methods* 18:284–304.

Gravlee, C. C., D. P. Kennedy, R. Godoy, and W. R. Leonard. 2009. Methods for collecting panel data: What can cultural anthropology learn from other disciplines. *Journal of Anthropological Research* 65:453–83.

Groves, R. M. 1987. Research on survey data quality. *The Public Opinion Quarterly* 51:S156–72.

Heerwegh, D. 2006. An investigation of the effect of lotteries on web survey response rates. *Field Methods* 18:205–20.

King, G., J. Honaker, A. Joseph, and K. Scheve. 2001. Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review* 95:49–69.

Laurie, H., R. Smith, and L. Scott. 1999. Strategies for reducing nonresponse in a longitudinal panel survey. *Journal of Official Statistics* 15:269–82.

McCarty, C., and P. Killworth. 2007. Impact of methods for reducing respondent on personal network structural measures. *Social Networks* 29:300–15.

Munger, G. F., and B. H. Loyd. 1988. The use of multiple matrix sampling for survey research. *Journal of Experimental Education* 56:187–91.

Neumayer, E. 2002. Do we trust the data? On the validity and reliability of cross national environmental surveys. *Social Science Quarterly* 83:332–40.

- Norris, F. H., C. K. Baker, A. D. Murphy, and K. Kaniasty. 2005. Social support mobilization and deterioration after Mexico's 1999 flood: Effects of context, gender, and time. *American Journal of Community Psychology* 36:15–28.
- Norris, F. H., and K. Kaniasty. 1996. Received and perceived social support in times of stress: A test of the social support deterioration deterrence model. *Journal of Personality and Social Psychology* 71:498–511.
- Norris, F. H., A. D. Murphy, C. K. Baker, J. L. Perilla, F. G. Rodriguez, and J. G. Rodriguez. 2003. Epidemiology of trauma and posttraumatic stress disorder in Mexico. *Journal of Abnormal Psychology* 112:646–56.
- Norris, F. H., J. L. Perilla, and A. D. Murphy. 2001. Postdisaster stress in the United States and Mexico: A cross-cultural test of the multicriterion conceptual model of posttraumatic stress disorder. *Journal of Abnormal Psychology* 110:553–63.
- Norris, F. H., D. L. Weishaar, M. L. Conrad, E. M. Diaz, A. D. Murphy, and G. E. Ibanez. 2001. A qualitative analysis of posttraumatic stress among Mexican victims of disaster. *Journal of Traumatic Stress* 14:741–56.
- Pasick, R. J., S. L. Stewart, J. A. Bird, and C. N. D'Onofrio. 2001. Quality of data in multiethnic health surveys. *Public Health Reports* 116:223–43.
- Pickery, J., G. Loosveldt, and A. Carton. 2001. The effects of interviewer and respondent characteristics on response behavior in panel surveys: A multilevel approach. *Sociological Methods & Research* 29:509–23.
- Raghunathan, T. E., and J. E. Grizzle. 1995. A split questionnaire survey design. *Journal of the American Statistical Association* 90:54–63.
- Sanchez, M. E. 1992. Effects of questionnaire design on the quality of survey data. *Public Opinion Quarterly* 56:206–17.
- Singer, E., J. Van Hoewyk, N. Gebler, T. Raghunathan, and K. McGonagle. 1999. The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics* 15:217–30.
- Storms, V., and G. Loosveldt. 2004. Who responds to incentives? *Field Methods* 16:414–21.
- Vandecasteele, L., and A. Debels. 2007. Attrition in panel data: The effectiveness of weighting. *European Sociological Review* 23:81–97.
- van den Brakel, J. A., R. Vis-Visschers, and J. J. G. Schmeets. 2006. An experiment with data collection modes and incentives in the Dutch Family and Fertility Survey for young Moroccans and Turks. *Field Methods* 18:321–34.

van der Zouwen, J., and T. van Tilburg. 2001. Reactivity in panel studies and its consequences for testing causal hypotheses. *Sociological Methods and Research* 30:35–56.

World Health Organization. 1997. *Composite International Diagnostic Inventory, Version 2.1*. Geneva: WHO.

Zeger, L. M., and N. Thomas. 1997. Efficient matrix sampling instruments for correlated latent traits: Examples from the National Assessment of Educational Progress. *Journal of the American Statistical Association* 92:416–25.