

SURVEYS AND SECONDARY DATA SOURCES**Using Survey Data in Social Science Research in Developing Countries**

Albert Park

The goal of this chapter is to introduce some of the major issues related to the use of survey data in social science research in developing countries. I will not address all aspects of survey research; nor will my treatment of the issues selected be comprehensive. Rather, my intention is to provide a broad roadmap of important issues to be addressed so that scholars embarking on field research in developing countries can think more systematically about (a) the potential value of quantitative data in research design, (b) how to find and use survey data collected by others, (c) when and how to conduct one's own survey, and (d) how to best utilize survey data in one's analysis. My insights are drawn from my own field research experiences as an economist, primarily in China, and come disproportionately from studies of economic development. Nonetheless, the main points are relevant to social scientists in other fields as well. One of my goals is to provide advice I wish I had been given at an earlier stage in my career, even though I am quite sure that many lessons can be learned only through firsthand experience.

Quantitative Data in Research Design

Ever since Charles Booth first attempted to measure poverty among the working class in England at the end of the 19th century (Booth 1889), there has developed a rich tradition of

quantitative data analysis to answer empirical questions in the social sciences. The great advantage of survey data is that they facilitate quantitative analysis that allows for generalization to an entire population. This ability to generalize relies on the very powerful properties of sampling and statistical theory. The epistemological basis of quantitative analysis contrasts quite sharply with that of qualitative research methods, such as ethnography, which rely more on inductive logic and the researcher's ability to synthesize a complex array of multiple sources of information. As a result, the choice of quantitative versus qualitative methods often depends on the research question. Surveys by their very nature are designed to reach a large number of respondents, but they provide only limited information on each unit of observation. They are thus often ill suited for the study of nuanced questions related to identity, subjective experience, or historical causation. But surveys are well suited to studies of political opinions; social attitudes; economic decisions and outcomes; and easily observable social behaviors such as fertility, migration, living arrangements, or visits to the doctor.

For many research questions, there can be great advantage to combining quantitative and qualitative approaches. In fact, these inherent complementarities underlie the value of interdisciplinary research and area studies. Thus, scholars for whom generalization is not a priority should not dismiss the possibility that survey data can be useful to their research. Similarly, those who consider generalization the primary objective of their work should not discount the importance of informing their analysis with less-structured interviews and observations.

How might the use of survey data complement qualitative studies? First, surveys can help frame questions or identify case sites for in-depth study. For example, if a researcher is interested in studying sites in both rich and poor regions, or in regions with both extended families and

nuclear families, survey or census data can point to appropriate locations. Often, one will want cases that are reasonably representative of a certain typology or that exhibit multiple characteristics (e.g., industrial and ethnically diverse communities). While informants can steer you in the right direction, quantitative evidence documenting whether a case is representative may be more convincing to many audiences. Depending on the number of cases, it might be appropriate to use quantitative data as the basis for randomized selection of cases to be studied in greater depth.

Survey data can also complement qualitative work by providing context for or confirming the findings of qualitative work. For example, if interviews reveal that rural women complain about greater work burdens on the farm when husbands move to urban areas for temporary work, it could be useful to elaborate that finding with data from a quantitative study of the extent of migration in the region or changes in female participation in agricultural work in migrant and nonmigrant households. Such data may be available in household surveys.

Quantitative data analysis is becoming increasingly accessible to social scientists across the disciplines as more and better data sets become available, as advances in computing speed and user-friendly software make it possible to learn how to analyze survey data with reasonable time investments, and as methods of analysis become increasingly powerful.

Census and Survey Data

A census is a complete enumeration of a population. It can be considered a particular type of sample survey in which the sampling rate is 100%! Many national governments carry out a population census once every 10 years. This is an enormously expensive undertaking, but it provides valuable if limited information on the entire population. The very large scale of a

census, and for that matter any large-scale survey, inevitably leads to problems of data quality. Certain populations, such as the illiterate or those who move frequently, may be difficult to enumerate, and the training of a large number of field staff may be difficult to regulate properly. Because they are undertaken infrequently, population censuses quickly become out of date for policy-making purposes. Of course, it is possible to conduct censuses on a smaller scale, for example by surveying all households in a village or all traders in a marketplace. However, for many research questions, social surveys that use probability sampling can provide a much more economical way to collect data since survey data enable one to make inferences from a small sample to a larger population.

Surveys may be conducted for a variety of reasons and by different types of individuals and organizations. Governments often carry out regular monitoring surveys to track outcomes of policy interest. For example, labor force surveys track changes in unemployment and other labor outcomes; income and expenditure surveys track changes in living standards, poverty, and inequality; and enterprise surveys monitor changes in investment and production. Government offices also keep administrative records of all types, usually based on reports made by lower-level government offices or other reporting units (e.g., tax payments, educational enrollment in schools, number of doctors and beds in hospitals). Such records can be a valuable resource for quantitative data analysis or for use as sampling frames for social surveys. However, both sampling error (a sample that is nonrepresentative of the population) and nonsampling error (reporting bias) may be major issues in interpreting administrative data.

In recent years, a large number of high-quality surveys have been conducted in developing countries throughout the world, and many of these data sources have become publicly available, often through the Internet. In fact, the first national general social survey in

the world was established in a developing country in 1950—the Indian National Sample Survey. This fact should help allay concern that surveys cannot be conducted reliably in developing-country settings. The “Secondary Data Sources” section of the Bibliography following this chapter provides information on some useful places to begin searching for publicly available survey data from developing countries. The Inter-university Consortium for Political and Social Research, founded in 1962, has a large archive of well-documented, publicly available data sets from more than 130 countries. In 1980, the World Bank established the Living Standards Measurement Study (LSMS) to explore ways of improving the type and quality of household data collected by government statistical offices in developing countries. LSMS surveys have been conducted in 30 developing countries, and the data are publicly available on a website managed by the Bank’s Development Research Group. Demographic and health survey data are available for 68 developing countries. These surveys build on a long history of active social surveys in the demography field (knowledge, attitudes, and practices (KAP) surveys of fertility in the 1960s, the World Fertility Survey and Contraceptive Prevalence Surveys in the 1970s and 1980s). The Roper Center for Public Opinion Research maintains a data archive of opinion research in more than 30 countries. Finally, research institutes such as the International Food Policy Research Institute and RAND have directed surveys in many developing countries. The U.S. government, through the National Science Foundation and the National Institutes of Health (NIH), has also supported surveys in developing countries. The National Institute of Child Health and Human Development, an institute under NIH, maintains a website providing access to surveys it has supported in eight developing countries.

Increasingly, surveys are conducted explicitly for research purposes. For example, two longitudinal household surveys that have been very influential in development economics are the

International Crop Research Institute for the Semi-Arid Tropics data set in India and the Indonesia Family Life Survey. Both have been the source of numerous research articles in the development literature. For low-budget surveys often conducted by academics, especially doctoral students, the feasible sample size may be much more limited. However, this limitation can be offset in many cases by the ability of the researcher to include original survey questions that address new and interesting questions. Many influential research papers across social science disciplines have been published using data sets with several hundred observations or fewer.

Using Survey Data Collected by Others

If one is interested in examining survey data for a specific country, one should actively search for existing data sources on the Internet and at national and local government offices, libraries and archives, research institutes and universities, or even enterprises or other community organizations. The discovery of such information can alter one's research design. But be prepared to abide by stipulations governing data use or to be asked to pay for the data.

If one finds that an existing data set may be valuable, it is helpful to ask questions about the data to assess its quality, possible biases, and appropriateness for one's research question. First, who collected the data and for what purpose? Sometimes, the announced goal of a survey can influence the type of responses elicited. For example, businesses are likely to underreport revenue and profits to the tax authority. An evaluation survey of a project conducted by the implementing organization could be biased toward producing a positive assessment. Second, what sampling method was employed? Third, what was the response rate? Fourth, how are key questions worded or key variables defined? If the definitions are vague or nonstandard, or if the measurement is prone to error, one might think twice about relying on such data. Fifth, how was

the survey conducted? Who developed the questionnaires, and how? Are the survey procedures well documented? How did training and supervision take place? Is there a substantial amount of missing data? Finally, who else has used the data? One can learn a great deal about the data by talking to other people who have used it.

If one decides to use survey data collected by others as a main part of one's research, it makes sense to invest in learning more about the data. It can be extremely informative to visit some of the surveyed areas and ask questions about topics that will be the focus of your study. This step can provide insights into local institutions, cultural practices, and social behaviors that may affect how you interpret the data. It is also a good idea to talk with researchers, government officials, or enumerators (the people who actually interviewed the respondents) involved in the survey. Ask them about their views of the strengths and weaknesses of the data. Sometimes you may learn that there were difficulties with the survey in some sites but not in others or difficulties with some sections of the questionnaire but not others. This is the type of information that one learns as a matter of course in running one's own survey but that can be obtained only through inquiry—and even then only to some extent—when using data collected by others.

Sometimes, extant data do not exist, or the questions you are trying to answer have not been asked in a survey. At this point, it may be important to consider conducting your own survey. In the next section, I briefly discuss various factors to consider when conducting a survey of your own design.

Conducting Surveys in Developing Countries

Should You Conduct Your Own Survey?

Conducting one's own survey in a developing country can be a major undertaking, and

the costs of such a project in terms of both time and money should not be underestimated. The decision to conduct a survey is a decision to become not just a scholar but also a project manager, a fundraiser, a survey methodologist, and a motivator and supervisor of others. On the other hand, because one's own survey can be tailored to specific research questions and yield data never before analyzed by others, it carries high potential to yield truly new insights. There probably is no better way to develop reliable intuition about what issues are empirically (not just theoretically!) meaningful than to spend extensive time in the field conducting one's own survey. The experience of conducting a survey also provides insights that help one better evaluate the quality of empirical research conducted by others. Working with local colleagues on surveys strengthens collaborative relationships that can form the basis of future work. From this perspective, conducting a survey can yield long-term benefits to one's research that go far beyond the actual data collected. For good overall reviews of issues related to fieldwork in developing countries, see Bulmer and Warwick (1993) and Devereux and Hoddinott (1992).

Collaboration

Nearly all successful surveys in developing countries depend on the support of energetic, capable research collaborators from the host country who know how to get things done within the country's institutional, political, and social environment; are skilled at interacting with government officials and community leaders; have developed reputations within the country that build trust; and have valuable substantive insights into the research questions. On the flip side, collaborators pursuing agendas at cross-purposes with those of the researcher can easily frustrate research plans.

In some cases, researchers, and especially junior researchers, have limited ability to control either the organization or the individuals with whom they must deal. In such cases, it

becomes critical to cultivate a positive working relationship. Part of this involves making sure that the collaborators perceive benefits to making the collaboration a success. The benefit bestowed can be intellectual—get them excited about the important contribution of the research; offer a course or lead a reading group on current research questions or methods; encourage collaborative research products; or help the collaborators realize their own intellectual goals (for example, through a visit to your institution). It can also be material—always provide adequate compensation for services, be as generous as possible to the host institution in the project budget. Finally, it might be personal—take the time to develop personal relationships, be helpful and friendly. In many cases, collaborators will be generous with their time and support, and it is important to reciprocate by sharing one’s research results and giving back in other ways, especially by building research capacity in the host institution. However, if for any reason you lose trust in those with whom you are working, be proactive in changing your collaborators even if the change is awkward or costly.

Combining Surveys with Natural Experiments and Randomized Interventions

Recently, there has been growing interest in combining social surveys with natural experiments or randomized interventions to enable more convincing causal inference. Natural experiments are unusual circumstances with a random element that facilitates comparisons. For example, a new national policy such as universal health care or a major infrastructure investment program might facilitate a comparison between individual behavior before and after the introduction of the program. Or a lottery to determine whether one is drafted, wins money, or can send children to a new charter school can be used to compare treatment and control groups that have similar characteristics. Randomized interventions can also be incorporated directly into research designs (Duflo & Kremer forthcoming). Early studies of randomized health insurance

policies (Manning et al. 1987) and randomized worker training programs (see Lalonde 1986 for a review) remain extremely influential. Recent examples in developing countries include the provision of education tuition subsidies for primary students in Mexico (Schultz 2004), the provision of school supplies such as flip charts and textbooks in rural Kenya (Kremer 2003), and the provision of iron supplementation to adults in Indonesia (Thomas et al. 2003). These projects are often partnerships between researchers and governments or nongovernmental organizations. Randomized interventions bring their own challenges. Members of control and treatment groups may react to the intervention in ways that are confounding (for example, individuals move from areas without interventions to areas with interventions), or the intervention may have spillover effects on the control group (for example, recipients of development project loans could have more funds to lend informally to nonrecipients).

Combining Surveys with Existing Surveys or Data Sets

If conducting one's own survey seems daunting or impractical, another model is to design a supplemental survey that can be combined with information from a survey that has already been completed or is ongoing. Whether this is a realistic option depends on whether one can find appropriate partner surveys and whether the project managers for those surveys are cooperative.

Joining forces often offers advantages to both sides. The most obvious is that with cost sharing, the sample size can be increased. Also, the researcher can benefit from the institutional capacity that has already been developed for the preexisting survey, while the manager of the partner survey may gain from the energy and expertise of the researcher. The disadvantage to the researcher is lost control over many key aspects of the survey, including timing, coverage, and content.

Because surveys are such major undertakings, it is wise to fully explore the possibility of

collaborating with others. Even government monitoring surveys are fair game because many government organizations are keen to earn extra revenues or may be interested in your research topic and so might consider allowing you to add supplemental questions to a subsample of their survey. If a local research organization has already completed a survey with questions you find interesting, you might work with researchers at that organization to resurvey the same sample. The advantages of resurveys are the saved costs from using data that have already been collected and the possibility of collecting longitudinal data, which follow respondents over time. One disadvantage may be difficulty locating all of the earlier respondents, leading to nonrepresentativeness caused by nonrandom attrition of the sample over time.

Keep in mind that collaboration is sure to involve many compromises. To ensure that such compromises are likely to be acceptable to both sides, each side should communicate its priorities and expectations clearly, especially regarding key survey-related issues such as sampling and monitoring. Researchers should be attentive to the critical importance of being able to verify the quality of the survey (i.e., sampling, training, data collection, and monitoring). Agreements should be put in writing.

Sampling

Sampling is one of the most critical aspects of any survey because it forms the basis for the key claim of generalizability, which is the main strength of quantitative research. Sampling designs can be very complicated, depending on the goal of the research (Kish 1965 is a classic reference). For example, researchers can stratify the sample to ensure sufficient variation in an independent variable of interest (e.g., first divide the population into income groups and then sample randomly within each group). Members of specific groups of interest (e.g., the poor) may be oversampled, or sampling may be based on an outcome of interest (e.g., participants and

nonparticipants in a program). The choice of sampling design often balances research goals and the costs of conducting the survey.

At the heart of sampling is the pursuit of representativeness through random selection. However, intuitions about sampling can be tricky. For example, one might randomly pick a set of villages in an area with equal probability and then randomly sample a fixed number of households in each village. But if poor villages are systematically smaller in population than nonpoor villages, such a sampling scheme will overrepresent the poor. There are three possible solutions to this problem: (a) make sure the probability that a village is chosen is proportional to its population, and sample a fixed number of households per village; (b) sample a fixed proportion rather than an absolute number of households in villages selected with equal probability; or (c) use the above approach and record the population of sampled villages to create sampling weights that could be used to correct the bias when using statistical methods later.

The first step in sampling is to clearly define the population of interest, such as rural households in the country or region or those 60 years of age and older who live in a household with children. The next step is to develop a sampling frame, or a list of possible respondents. First, one should seek out appropriate census data or administrative records (e.g., village household registers). Be aware that outdated lists will underrepresent newly formed households or individuals. In many cases, unfortunately, preexisting lists are unavailable or require updating. In China, for example, rural migrants living in cities are notoriously difficult to study because they have left the countryside, are highly mobile, and often do not bother registering with local community offices in cities. In such cases, one must expend resources to create a sampling frame from scratch, that is, send enumerators to all households in the neighborhood or settle for a population slightly different from that of theoretical interest, such as registered migrants rather

than all migrants.

A common approach to sampling is to first divide the population into geographical sampling units, such as villages. In a first stage, sample villages are selected based on existing village-level information about population, income, and so on. An individual sampling frame is then constructed only in the sampled villages. Statistics using data from multilevel, or clustered, sampling must adjust for the likely correlation in characteristics of observations drawn from the same cluster. Some surveys have also adopted spatial sampling approaches, often with the assistance of aerial or satellite maps. The unit of observation becomes a dwelling, and enumerators visit dwellings and interview whomever they find there. An application of this method using GPS technology in urban China, where unregistered migrant workers are numerous, is described in Landry and Shen (2005). GPS technology is also making it possible to collect, at reasonable costs, geographic data such as location, elevation, land size, and so on, as part of household village surveys.

Armed with a sampling frame, one can randomly select respondents to be surveyed. However, there is no guarantee that all of the selected individuals can be found and will participate in the survey. Nonresponse can lead to an unrepresentative sample. Appropriate timing of visits can help increase the response rate (e.g., visit during evening hours if individuals work during the day, or visit on holidays, when household members are normally at home). Introductions by village or neighborhood leaders can help reduce refusal rates. A protocol should be developed to deal with nonresponse. Enumerators should be required to visit a respondent on several different days before giving up. If the respondent is replaced, the replacement should be randomly selected from a backup list sampled at the same time as the original sample. If the sampling frame provides information on the characteristics of the individual or household, one

can choose replacements that resemble the original household, although this approach does not eliminate the possibility of bias.

Researchers must consider trade-offs among sampling error, nonsampling error, and cost in making decisions in the field regarding sample selection. For example, very remote villages may be inaccessible by vehicles, thereby substantially increasing survey costs. Similarly, some individuals may be illiterate and unable to answer the questions in the survey reliably or in a reasonable period of time.

Another important aspect of sampling is the choice of an appropriate sample size. Statistically, a larger sample makes it possible to measure the relationships between variables with greater precision. More formally, the outcome of statistical inference has four components: sample size; effect size; significance level (required statistical confidence level, usually set at 95%); and power, or the odds that one will observe a treatment effect if it occurs. The minimum sample size necessary for precise estimation depends on the strength of the empirical relationship, which sometimes can be roughly approximated *ex ante* by making use of results from other studies. A useful reference for evaluating the power of statistical tests given a chosen sample size is Cohen (1988). In my experience, sample sizes of fewer than 200 observations often make it challenging to employ appropriate empirical methods or to produce results that are statistically significant.

Questionnaire Design

Clearly defining the research question is the first step in designing an effective questionnaire. One common mistake made by researchers interested in a specific issue area is to ask an exhaustive set of questions about everything that is important, unusual, or interesting about the issue without having a clear research question or methodology in mind. The result

often is a long and unwieldy questionnaire that still omits critical information. Survey methodologists have shown that the quality of information that survey respondents provide declines significantly after more than 30 minutes, and the length of the questionnaire also dictates the amount of time enumerators must spend on each interview. Thus, there are costs in terms of both data quality and money if survey questionnaires are too long. Interesting but inessential questions should be eliminated.

In designing questionnaires, do not try to reinvent the wheel. Many smart people have spent a lot of time refining questions to study different topics, taking many factors into consideration. You should spend time finding out what other surveys have been conducted on the same or related issues, both in your country of interest and elsewhere. Ask for copies of questionnaires; most researchers or government officials are happy to provide copies of blank questionnaires. There are advantages in making questions comparable to surveys done elsewhere or to official statistical conventions used in the country; doing so can facilitate comparisons that will allow you to better contextualize or generalize your results. A particularly valuable resource for designing questionnaires for household surveys in developing countries is the edited volume by Grosh and Glewwe (2000), in which experienced experts summarize the current state of the art in questionnaire design for various topics (e.g., income and expenditure, agricultural production, self-employment, rural credit, education, health, etc.).

It goes without saying that questions included in survey instruments should be worded in culturally appropriate language. Active participation by local collaborators, informants, and study subjects is invaluable. Questions should be simple, direct, and familiar to respondents. Each question should be stated in a neutral, not a leading, manner and whenever possible should be written out completely to avoid misinterpretation. Because it is difficult to know what

circumstances respondents will attribute to a hypothesized situation, hypothetical questions should be resisted. Two common criteria used to evaluate a question are reliability (whether the question produces a consistent response) and validity (whether the question is actually measuring what it was intended to measure). The former can be tested through pretesting, and one check of the latter is to translate and retranslate the questions. It takes a surprisingly long period of time and quite a few iterations to design an effective questionnaire. Some references that discuss more specific aspects of questionnaire design include Fowler (1995); Warwick and Lininger (1975); and Brislin, Lonner, and Thorndike (1973).

One common pitfall in questionnaire design is inadequate budgeting of time and resources for presurvey work, which often starts with open-ended interviews and ends in pretests of survey instruments to be used in the main survey. Often, time and resources to complete the project are limited, and researchers want to save resources to maximize the amount of data they can obtain from the main survey. However, the quality of the data is no better than the questionnaire, and many serious problems can arise if you don't do adequate pretesting in multiple survey sites. Seemingly straightforward questions may be interpreted by respondents in ways the researcher did not intend or predict.

Fielding a Survey and Managing the Data

Enumerators are the individuals responsible for interviewing respondents and so play a critical role in determining the quality of survey data. They can be university students (undergraduates or graduate students), research institute staff, or professional survey staff (e.g., those who work for a government statistical agency or other established survey organization). An advantage of using students is that they are usually bright, energetic, willing to follow the directions of supervisors, and relatively cheap to hire. The disadvantage is that they lack survey

experience, are less socially mature, and may lose interest if the work is unusually difficult or tiring. The advantage of professional staff is that they are experienced, understand local circumstances and so are less likely to misinterpret questions, and have well-established relationships in local communities. The disadvantage is that they may be inflexible about their work procedures and motivated primarily by financial gain.

Training of enumerators should include basic interview techniques (e.g., polite, respectful behavior and attitude; how to avoid asking questions in a leading way); explanation of the goals of the research; and discussion of survey protocols, the meaning and purpose of each survey question, and protocols to follow when respondents do not understand questions or refuse to answer. A good training program should provide written materials that summarize key points and are easy to read (to serve as a reference for enumerators), include extensive practice interviews, and require that enumerators demonstrate competence through a formal evaluation exercise. Videotapes of positive and negative examples of enumerator behavior can serve as an effective teaching tool. When the survey starts, less experienced enumerators and experienced enumerators should visit households together for the first several interviews.

It is critical that the work of enumerators be monitored systematically. In surveys that I have directed, supervisors review completed questionnaires each evening and discuss problems with enumerators before the next day's work. After several days, the frequency of mistakes declines substantially. One can also ask enumerators to check each other's questionnaires. A written set of standard internal consistency checks (e.g., components add up to their sum, labor activities match up with reported income sources) should be provided to enumerators and be the basis of supervision checks. If the survey is occurring in several places at once, frequent communication during the earlier stages of the survey can ensure that issues that arise in the field

can be addressed in a consistent manner.

It is important to revisit a subset of respondents to verify that enumerators actually interviewed respondents. If households have phones, such checks can be done by telephone. Usually, asking just a few questions from different parts of the survey instrument is sufficient to complete a check. Enumerators should be required to revisit any households with substantial discrepancies, and if one enumerator has multiple problem surveys, the enumerator should be fired, and all of the households visited by that enumerator should be revisited by another person. Ideally, some monitoring should be done by individuals from an organization different from that of the enumerators. Both enumerators and their supervisors should be given incentives in the form of praise, bonuses, or prizes.

Once survey questionnaires have been completed, a number of decisions must be made about data entry and cleaning. Because entering data into computers is labor intensive, it is often cost effective to enter the data while in the host country. One can usually find staff of research institutes, university students, or private companies to do the work. The price should always be based on work completed (number of forms) rather than time spent. The contract should include financial incentives for fast completion, be based on quality standards evaluated by a method agreed on in advance, and stipulate final payment only after all of the work has been completed.

If the questionnaire is short and relatively straightforward, spreadsheet software such as Excel can be used for data entry. However, in general, one will want to use a database program for which entry screens can be designed to look exactly like the questionnaires themselves (e.g., Access, dBase). The software should make it possible to enter range restrictions, such as not allowing months to be greater than 12, and internal consistency checks, such as requiring that children's age be less than parents' age or that total expenditures equal the sum of the

components of expenditure. It takes some investment of time to learn the software and create the database structure and entry screens, so it may be a good idea to budget resources for an experienced programmer, if necessary. In developing countries, computers may have limited capacity, and recent versions of software requiring large memory may run slowly or not at all. I have had great success using an old version of a simple software program based on a disk operating system. The software provides essential functions but can run easily on even primitive computers.

If possible, data entry should occur in a fixed location with a supervisor familiar with the questionnaire always present to troubleshoot questions, such as how to deal with specific coding problems or use the software. Special codes should be used for different types of missing data (e.g., the respondent did not know the answer, the respondent refused to answer, the question was not applicable) since these distinctions can affect interpretation of results. If possible, data from each form should be entered twice, preferably by different individuals, and the two entries should be compared (this usually requires some programming). If double data entry is too costly, one must at least design an entry-checking protocol for manually checking the entered data against the questionnaires. Hard copies of questionnaires should be archived for future reference.

Some surveys now enter the data in the field, by means of notebook computers, so that problems can be discovered almost immediately and enumerators can make return visits to households without delay. While this practice can reduce the cost of revisits and improve data quality, it requires greater advance preparation and can increase costs because of greater hardware requirements and the need to support living expenses of data entry staff in the field. A usually prohibitively costly option is to equip enumerators themselves with notebook computers and bypass paper questionnaires altogether.

Analyzing Survey Data

Advances in statistical software, both in user friendliness and in the capacity to conduct various types of statistical analysis, make it increasingly easy to analyze empirical data with just a modest investment in learning new software. Among empirical economists and many other social scientists, Stata has become by far the most popular software for data management and statistical analysis, while SPSS (originally known as Statistical Package for the Social Sciences) remains a popular choice among sociologists and political scientists. These software packages now include excellent tutorials that make it relatively easy to summarize data (i.e., calculate sample and subsample means), produce tables and figures, and run regressions.

A good introduction to recent empirical methodologies for analyzing household surveys in developing countries is Deaton (1997). Another useful reference is Sadoulet and de Janvry (1995). Space constraints preclude a detailed discussion of the statistical analysis of survey data, but I will point out a few useful but frequently overlooked analytical tools that are easily implemented in many software packages:

- ❖ In reporting differences in the mean values of variables for two different groups, it is usually a good idea to test whether the reported differences are statistically significantly different from zero.
- ❖ The sampling design of the survey may require that sampling weights be used in statistical calculations. Multistage clustered sampling designs usually require that standard deviations and standard errors of regression coefficient estimates be adjusted for likely within-cluster correlations (Deaton 1997).
- ❖ Recently popularized nonparametric methods can produce revealing visual pictures of the distribution of a variable or the bivariate relationship between two variables that

allow the slope relationship to vary over different parts of the distribution (Deaton 1997; Pagan & Ullah 1999).

❖ One common problem is estimating how x affects y even though y may also affect x , or some third factor may affect both x and y . Panel data (multiple respondents observed over multiple periods) can facilitate over-time comparisons that control for bias from many omitted variables (Hsiao 1986; Wooldridge 2002). If such data are combined with a natural experiment or a randomized intervention, one can compare the change that occurs in the treatment group with the change that occurs in the control group, the so-called difference-in-difference estimator (Wooldridge 2002). Another common statistical approach to identifying the effect of x on y is to use instrumental variables, which are variables that strongly affect x but do not affect y independently of x (for a clear explanation of this method, see Wooldridge 1999).

Final Thoughts on Research Involving Survey Data

Defining a research question is an iterative process. It often starts with a researcher's initial interest in a topic, stimulated by personal experience; the interest and encouragement of an adviser, mentor, or colleague; close reading of the existing literature in a particular discipline; or knowledge of an issue or circumstance in a specific region of the world. One useful question to ask oneself in choosing among potential research questions to be studied in a specific country or region is, "Why is this question of particular interest in this country or region?" This helps avoid a common mistake of mechanically deciding to address a currently popular question in the country of one's area expertise.

Once a question is chosen, the process of field research almost invariably alters the way in which the question is framed or posed, as well as the research design. All social scientists who

have engaged in extensive field research in developing countries have a story to tell about how the substantive focus of a research project changed dramatically after they spent time in the field. Iteration frequently occurs when field research presents new insights into the question or reveals new opportunities for collecting information (e.g., the unexpected discovery of an archive; a useful informant or contact; a natural experiment; or an interesting social behavior, policy, or outcome). This circumstance may cause a rethinking of the research question itself or of the strategy for studying the question.

Many unexpected things can occur during field research, such as delays due to political events, natural disasters, institutional factors, and so on. The demands placed on you may change (e.g., demands for more money, restrictions on research activity), or you may find that your interests shift or your original idea proves infeasible. It is a good idea to think about contingency plans for accomplishing your research goals well ahead of time, just in case things do not go as planned.

In the old days, when costly transport and communications isolated the researcher in the field for months at a time, many of these judgments had to be made alone. In the new world of cheaper air flights, long-distance calling cards, and the Internet, it is possible for researchers to iterate between field research and academic reflection and consultation while residing at their home institution or in the field. Such an approach is strongly recommended because it is easy for researchers to lose perspective when fully immersed in field research. Udry (2003) argues that surveys themselves can be designed to incorporate iterative learning and that such a design, while costly in time, can allow survey researchers to be more creative and more responsive to surprises.

The goal of this chapter has been to provide a general introduction to the use of survey

and census data in social science research in developing countries. I have argued that the collection or use of such data can be a useful component of many field research plans and research designs, whether one plans to rely primarily on quantitative or qualitative methods. The increasing availability of data sets and the rapid advances in the ability of computing software to manage and analyze data make it more feasible than ever before to use survey and census data in one's research.

However, conducting surveys in developing countries can be challenging, and numerous factors at every stage of the survey process, from sampling to questionnaire design to data entry, can affect the quality of the data and so the quality of the research. Much of this chapter has been spent pointing out potential pitfalls in the collection of one's own data and issuing warnings about quality issues in survey data produced by others. I hope that social scientists across the disciplines will feel confident in using survey data in their research and will find such data a valuable component of their research methodology.

Supplemental References

Booth, Charles. (1889). *Labour and Life of the People*. London, UK: Macmillan.

Brislin, Richard, Walter J. Lonner, and Robert M. Thorndike. (1973). *Cross-Cultural Research Methods*. New York: John Wiley.

Duflo, Esther and Michael Kremer. (Forthcoming). "Use of Randomization in the Evaluation of Development Effectiveness." In *Proceedings of the Conference on Evaluating Development Effectiveness, July 15–16, 2003*. Washington, DC: World Bank Operations Evaluation Department.

Kremer, Michael. (2003). "Randomized Evaluations of Educational Programs in

Developing Countries: Some Lessons.” *American Economic Review* 93(2), 102–106.

LaLonde, Robert. (1986). “Evaluating the Econometric Evaluations of Training Programs with Experimental Data.” *American Economic Review* 76(4), 604–620.

Manning, Willard, Joseph Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz. (1987). “Health Insurance and the Demand for Medical Care: Evidence From Randomized Experiment.” *American Economic Review* 77(3), 251–277.

Schultz, T. Paul. (2004). “School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program.” *Journal of Development Economics* 74(1), 199–250.

Thomas, Duncan, Elizabeth Frankenberg, Jed Friedman, Jean Pierre Habichtl, Mohammed Hakimi, Nathan Jones, Jaswadi, Gretel Pelto, Bondan Sikoki, Teresa Seeman, James P. Smith, Cecep Sumantri, Wayan Suriastini, and Siswanto Wilopo. (2003). “Iron Deficiency and the Well-Being of Older Adults: Early Results From a Randomized Nutrition Intervention.” Unpublished paper.

Udry, Christopher. (2003). “Fieldwork, Economic Theory and Research on Institutions in Developing Countries.” *American Economic Review* 93(2), 107–111.

Bibliography: Surveys and Secondary Data Sources

1. Overview and Essentials

Alreck, Pamela and Robert Settle. (1995). *The Survey Research Handbook: Guidelines and Strategies for Conducting a Survey*. Chicago, IL: Irwin Press.

Babbie, Earl. (1990). *Survey Research Methods*. Belmont, CA: Wadsworth.

Bulmer, Martin and Donald Warwick. (1993). *Social Research in Developing Countries:*

Surveys and Censuses in the Third World. London, UK: University College London Press Limited.

Caldwell, John C. (1985). "Strengths and Limitations of the Survey Approach for Measuring and Understanding Fertility Change." In *Reproductive Change in Developing Countries*, edited by John Cleland and John Hobcraft. Oxford, UK: Oxford University Press.

Devereux, Stephen and John Hoddinott, eds. (1992). *Fieldwork in Developing Countries*. New York: Harvester Wheatsheaf.

Fink, Arlene. (1993). *Evaluation Fundamentals: Guiding Health Programs, Research and Policy*. Newbury Park, CA: Sage.

-----, ed. (1995). *The Survey Kit Series*. Thousand Oaks, CA: Sage.

Freedman, Ronald. (1987). "The Contribution of Social Science Research to Population Policy and Family Planning Program Effectiveness." *Studies in Family Planning* 18(2), 57–82.

Hess, Jennifer, Jennifer Rothgeb, and Andy Zukerberg. (1998). "Developing the Survey of Program Dynamics Survey Instruments." Statistical Research Division Working Papers in Survey Methodology, no. SM98/07. Retrieved September 8, 2005, from www.census.gov/srd/www/byyear.html

Hess, Jennifer and Eleanor Singer. (1995). "The Role of Respondent Debriefing Questions in Questionnaire Development." Statistical Research Division Working Papers in Survey Methodology, no. SM95/18. Washington, DC: U.S. Bureau of the Census. Retrieved September 8, 2005, from www.census.gov/srd/www/byyear.html

Grosh, Margaret and Paul Glewwe. (2000). *Designing Household Survey Questionnaires for Developing Countries: Lessons From 15 Years of the Living Standards Measurement Study*, 3

vols. Washington DC: World Bank.

Michigan Program in Research Methodology, Institute for Social Research at the University of Michigan, accessed September 8, 2005, at www.isr.umich.edu/gradprogram/

Oksenberg, Lois, Charles Cannell, and Graham Kalton. (1991). "Strategies for Pretesting New Questions." *Journal of Official Statistics* 7(3), 349–365.

Rea, Louis M. and Richard A. Parker. (1997). *Designing and Conducting Survey Research: A Comprehensive Guide*. San Francisco, CA: Jossey-Bass. See Chapter 1 ("An Overview of the Sample Survey Process"), Chapter 2, Chapter 3, Chapter 4 ("Administering the Questionnaire," in particular, sections on "Precoding the Survey Instrument" and "Editing the Completed Questionnaire"), Chapter 5, Chapter 6, Chapter 7, and Chapter 8.

Rossi, Peter, James D. Wright, and Andy B. Anderson. (1983). *Handbook of Survey Research*. Orlando, FL: Academic Press.

Schutt, Russell L. (1998). *Investigating the Social World*. Thousand Oaks, CA: Pine Forge.

Sudman, Seymour and Norman M. Bradburn. (1982). *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco, CA: Jossey-Bass. See Chapter 2, "Asking Non-Threatening Questions About Behavior"; Chapter 3, "Asking Threatening Questions About Behavior"; Chapter 4, "Questions for Measuring Knowledge"; Chapter 5, "Measuring Attitudes: Formulating Questions"; Chapter 6, "Measuring Attitudes: Recording Responses"; Chapter 7, "Using Standard Demographic Terms" (skim); Chapter 8, "Order of the Questionnaire"; Chapter 9, "Format of the Questionnaire"; and Chapter 10, "Designing Questions for Mail and Telephone Surveys."

Warwick, Donald P. and Charles A. Lininger. (1975). *The Sample Survey: Theory and Practice*. New York: McGraw-Hill.

Weisberg, Herbert and Bruce D. Bowen. (1977). *An Introduction to Survey Research and Data Analysis*. San Francisco, CA: W. H. Freeman.

2. Planning, Proposing, and Sampling for Survey Data

Cohen, Jacob. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: L. Erlbaum Associates.

Grosh, Margaret and Paul Glewwe. (2000). *Designing Household Survey Questionnaires for Developing Countries: Lessons From 15 Years of the Living Standards Measurement Study*, 3 vols. Washington DC: World Bank. See associated website: www.worldbank.org/lsms/index.htm

Kish, Leslie. (1965). *Survey Sampling*. New York: John Wiley.

Krathwohl, David. (1988). *How to Prepare a Research Proposal: Guidelines for Funding and Dissertations in the Social and Behavioral Sciences*. Syracuse, NY: Syracuse University Press. See Chapter 8, “A Checklist for Critiquing Proposals.”

Landry, Pierre, and Mingming Shen. (2005). “Reaching Migrants in Survey Research: the Use of the Global Positioning System to Reduce Coverage Bias in China, Political Analysis 13(1), Winter.

Ross, Kenneth N. (1992). “Sample Design Procedures for a National Survey of Primary Schools in Zimbabwe.” In *Issues and Methodologies in Educational Development no. 8*. [**Au: Please provide location of publisher.**]International Institute for Educational Planning, UNESCO.

Singleton, Royce A., Bruce C. Straits, and Margaret Miller Straits. (1993). *Approaches to Social Research*. New York: Oxford University Press. See Chapter 6, “Sampling.”

3. Qualitative Inquiry in and Beyond the Design Process

Bunte, Pamela A., Rebecca M. Joseph, and Peter Wobus. (1992). “The Cambodian Community of Long Beach: An Ethnographic Analysis of Factors Leading to Census Undercount.” United Cambodian Community, Inc., and Center for Survey Methods Research, Bureau of the Census. Ethnographic Evaluation of the 1990 Decennial Census Report no. 9 (March). Retrieved as report no. EV92/09 on September 9, 2005, from www.census.gov/srd/www/byyear.html

de Vaus, David A. (1996). *Surveys in Social Research*. London, UK: University College London Press. See Chapter 1, “The Nature of Surveys”; Chapter 2, “Theory and Social Research”; Chapter 3, “Formulating and Clarifying Research Questions”; and Chapter 4, “Developing Indicators for Concepts.”

Fuller, Theodore, John Edwards, Sairudee Vorakitphokatorn, and Santhat Sermsri. (1993). “Using Focus Groups to Adapt Survey Instruments to New Populations: Evidence From a Developing Country.” In *Successful Focus Groups*, edited by David L. Morgan. Newbury Park, CA: Sage.

Massey, Douglas. (1987). “The Ethnosurvey in Theory and Practice.” *International Migration Review* 21(4), 1498–1521.

Newby, Margaret, Sajeda Amin, Ian Diamond, and Ruchira T. Naved. (1998). “Survey Experience Among Women in Bangladesh.” *American Behavioral Scientist* 42(2), 252–275.

Rynearson, Ann M. and Thomas A. Gosebrink[**Au: Website below shows a third**

author (Leslie A. Brownrigg); OK to add?] (with Barrie M. Gewanter). (1990). "Barriers to Censusing Southeast Asian Refugees." *Bureau of the Census Ethnographic Exploratory Report no.10* (no. EX90/10). Report is 92 pages; retrieved September 9, 2005, from www.census.gov/srd/www/byyear.html

Trochim, William M. (2000). "Sampling." In *The Research Methods Knowledge Base*, 2d ed. Cincinnati, OH: Atomic Dog Publishing.

----- (2002). "Language of Research." In *The Research Methods Knowledge Base*, 2d ed. Available at www.socialresearchmehtods.net/kb/language.htm.

4. Developing Good Questions: Fundamental Concepts

Converse, Jean and Stanley Presser. (1986). *Survey Questions: Handcrafting the Standardized Questionnaire*, Series No. 07-063. Beverly Hills: Sage.

Fowler, Floyd. (1995). *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage.

Freedman, Deborah, Arland Thornton, Donald Cambum, Duane Aiwin, and Linda Young-DeMarco. (1988). "The Life History Calendar: A Technique for Collecting Retrospective Data." *Sociological Methodology* 18, 37-68.

Kalton, Graham. (1983). "Introduction to Survey Sampling." *Sage University Paper Series on Quantitative Applications in the Social Sciences*, Series No. 07-035. Beverly Hills, CA: Sage.

Sudman, Seymour and Norman M. Bradburn. (1982). *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco, CA: Jossey-Bass. See Chapter 1, "The Social

Context of Question Asking,” and Chapter 11, “Questionnaire from Start to Finish.”

Sweet, James, Larry Bumpass, and Vaughn Call. (1988). “The Design and Content of the National Survey of Families and Households.” *University of Wisconsin Center for Demography and Ecology NSFH Working Paper no. 1*. Madison: University of Wisconsin. Retrieved September 9, 2005, from www.ssc.wisc.edu/cde/nsfhwp/

5. Pretesting: Rationale and Overview of Field Techniques

DeMaio, Theresa, Jennifer Rothgeb, and Jennifer Hess. (1998). “Improving Survey Quality Through Pretesting.” *Statistical Research Division Working Papers in Survey Methodology*, no. 98/03. Washington, DC: U.S. Bureau of the Census. Retrieved September 9, 2005, from www.census.gov/srd/papers/pdf/sm98-03.pdf

Foddy, William. (1998). “An Empirical Evaluation of In-Depth Probes Used to Pretest Survey Questions.” *Sociological Methods and Research* 27(1), 103–134.

Schechter, Susan, Johnny Blair, and Janet Vande Hey. (1996). “Conducting Cognitive Interviews to Test Self-Administered and Telephone Surveys: Which Methods Should We Use?” *University of Maryland Survey Research Center Working Paper*.

6. Data Precoding, Coding, Cleaning, and Management

de Vaus, David A. (1996). *Surveys in Social Research*. London, UK: University College London Press. See Chapter 14, “Coding.”

Ross, Kenneth, T. Neville Postlethwaite, Marlaine Lockheed, Aletta Grisay, and Gabriel Carceles Breis. (1990). “Improving Data Collection, Preparation and Analysis Procedures: A Review of Technical Issues.” In *Planning the Quality of Education: The Collection and Use of*

Data for Informed Decision-Making, edited by Kenneth Ross and Lars Mahlck. Paris, France: UNESCO.

University of Michigan Survey Research Center. (N.d.). "Introduction to the Coding Procedures Used in the Survey Research Center." [Au: **Please provide publication information or a website so that readers can find this source.**]

7. Administration

Biemer, Paul P., Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, eds. (1991). *Measurement Errors in Surveys*. New York: Wiley.

de Vaus, David A. (1996). *Surveys in Social Research*. London, UK: University College London Press. See Chapter 7, "Administering Questionnaires."

Jenkins, Cleo R. and Don Dillman. (1995). "Towards a Theory of Self-Administered Questionnaire Design." Statistical Research Division Working Papers in Survey Methodology, no. SM95/06. Washington, DC: US Bureau of the Census. Retrieved September 9, 2005, from www.census.gov/srd/www/byyear.html

Lyberg, Lars E. and Daniel Kasprzyk. (1991). "Data Collection Methods and Measurement Error: An Overview." In *Measurement Errors in Surveys*, edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman. New York: Wiley.

Weeks, Michael. (1992). "Computer-Assisted Survey Information Collection: A Review of CASIC Methods and Their Implications for Survey Operations." *Journal of Official Statistics* 8(4), 445–465.

8. Fieldwork in Principle and Practice

Billiet, Jacques and Geert Loosveldt. (1988). "Improvement of the Quality of Responses to Factual Survey Questions by Interviewer Training." *Public Opinion Quarterly* 52(2), 190–211.

Current Population Survey Interviewers Manual. US Bureau of the Census. Retrieved September 9, 2005, from www.bls.census.gov/cps/bintman.htm. An example of an interview, examples of training, and interview manuals.

Fowler, Floyd J. (1993). *Survey Research Methods*. Newbury Park: Sage. See Chapter 3, "Nonresponse: Implementing a Sample Design."

Frey, James H. and Sabine Mertens Oishi. (1995). *The Survey Kit Volume 4: How to Conduct Interviews by Telephone and in Person*. Thousand Oaks, CA: Sage. See Chapter 3, "Interviewer Selection and Training."

Hua, Haiyan et al. **[Au: Please list all authors.]** (1997). "Field Supervisor's/Enumerator's Manual." In *Girls' and Women's Education Project Documentation*. Katmandu: World Education Nepal. Examples of training and interview manuals.

----- (1997). "GWE/Nepal Literacy Study Hypotheses and Indicators." In *Girls' and Women's Education Project Documentation*. Katmandu: World Education Nepal.

----- (1997). "GWE Survey: Program and Site Selection Information." In *Girls' and Women's Education Project Documentation*. Katmandu: World Education Nepal.

Igoe, Lin Moody, ed. (1993). *China Economic, Population, Nutrition and Health Survey 1993 Work Manual*. Beijing, China, and Chapel Hill, NC: Chinese Academy of Preventive Medicine and Carolina Population Center.

Examples

Rea, Louis M. and Richard A. Parker. (1997). *Designing and Conducting Survey Research: A Comprehensive Guide*. San Francisco, CA: Jossey-Bass. See Chapter 4, “Administering the Questionnaire.”

Watkins, Susan Cotts (with Naomi Rutenberg, Steve Green, Charles Onoko, Kevin White, Nadra Franklin, and Sam Clark). (1995). “‘Circle No Bicycle’: Fieldwork in Nyanza Province, Kenya, 1994–1995.” University of Pennsylvania Population Studies Center, Social Networks Project. Retrieved September 9, 2005, from www.ssc.upenn.edu/Social_Networks/Level%203/Papers/PDF-files/Circle-No-Bicycle.doc

Weinberg, Eva. (1983). “Data Collection: Planning and Management.” In *Handbook of Survey Research*, edited by Peter H. Rossi, James D. Wright, and Andy B. Anderson. New York: Academic Press, Inc. Focus on pp. 350–358.

9. Analysis of Survey Data

Deaton, Angus. (1997). *Analysis of Household Surveys*. Baltimore, MD: Johns Hopkins University Press for the World Bank.

Hsiao, Cheng. (1986). *Analysis of Panel Data*. Cambridge, UK: Cambridge University Press.

Pagan, Adrian and Aman Ullah. (1999). *Nonparametric Econometrics*. Cambridge, UK: Cambridge University Press.

Sadoulet, Elisabeth and Alain de Janvry. (1995). *Quantitative Development Policy Analysis*. Baltimore, MD: Johns Hopkins University Press.

Wooldridge, Jeffrey M. (1999). *Introductory Econometrics: A Modern Approach*. Cincinnati, OH: Southwestern College Publishing.

----- (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: Massachusetts Institute of Technology Press.

10. Secondary Data Sources

International Labor Office

International Labor Office: <http://laborsta.ilo.org/>

United Nations

United Nations Development Programs. [Annual]. *Human Development Report*. New York: Oxford University Press, various issues, also see www.hdr.undp.org/

World Bank

World Bank. [Annual]. *World Development Report*. New York: Oxford University Press.

World Bank Data and Statistics Division, accessed September 9, 2005, at www.worldbank.org/data/

World Bank, Living Standards Measurement Survey, conducted throughout the world; accessed September 9, 2005, at www.worldbank.org/html/prdph/lsms/

Some Recent Surveys Available Online

African Ideational Diffusion Project, accessed September 9, 2005, at www.acap.upenn.edu/

Demographic and Health Surveys, accessed September 9, 2005, at
www.measuredhs.com/

Indonesia Family Life Survey (IFLS) longitudinal survey of 30,000 individuals living in 13 of the 27 provinces in the country, RAND-directed surveys in 1993/94, 1997, 1998, and 2000, accessed September 9, 2005, at www.rand.org/labor/FLS/IFLS/

International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) Village Level Study Panel Data: accessed September 25, 2005 at www.econ.yale.edu/~egcenter/special.htm.
[available through the Yale Economic Growth Center Center for Data Sharing by request].

International Food Policy Research Institute, accessed September 9, 2005, at
www.ifpri.org/

Inter-university Consortium for Political and Social Research, accessed September 9, 2005, at www.icpsr.umich.edu/

Latin American Migration Project, accessed September 9, 2005, at
lamp.opr.princeton.edu/

Mexican Migration Project, accessed September 9, 2005, at
www.pop.upenn.edu/mexmig/

Nang Rong Surveys, accessed September 9, 2005, at
www.cpc.unc.edu/projects/nangrong/

National Institute of Child Health and Human Development, National Institutes of Health, accessed September 9, 2005, at www.nichd.nih.gov/about/cpr/dbs/res_ss_large.htm

RAND, accessed September 9, 2005, at www.rand.org/services/databases.html

RAND Family Life Surveys from Indonesia, Malaysia, Bangladesh, and Guatemala,
accessed September 9, 2005, at www.rand.org/labor/FLS/

Roper Center for Public Opinion Research, accessed September 9, 2005, at
www.ropercenter.uconn.edu/

South African Household and Livelihood Survey, accessed September 9, 2005, at
www.worldbank.org/html/prdph/lsms/country/za94/za94home.html

World Fertility Surveys, accessed September 9, 2005, at opr.princeton.edu/archive/wfs/