(Extended Abstract)

Comparison of data quality between conventional and continuous Demographic and Health Surveys in Peru

Yoonjoung Choi, Rachel Lucas, Madeleine Short, Nisha Sarpal, Jacob Adetunji
US Agency for International Development

Background

The Demographic and Health Surveys (DHS) are nationally representative household surveys providing data on health and population in developing countries. The DHS project, supported by the US Agency for International Development (USAID), has provided technical assistance to more than 260 surveys in over 80 countries since 1984. Most countries have implemented a series of surveys for planning and monitoring and evaluation of population, health, and nutrition programs, typically in about every 5 years.

In Peru, four DHS were conducted between 1986 and 2000. However, with strong demand for more frequent monitoring of health and population indicators, the DHS project used the continuous survey methodology – a survey method involving rounds of data collection at regular intervals to allow an ongoing assessment of indicators. While the method has been used widely in developed countries (e.g., the Consumer Expenditure Survey and the National Health and Nutrition Examination Survey in the United States), the use of the method was experimented for the first time in the DHS project in Peru. The first set of Peru Continuous DHS (CDHS) included 5 annual rounds of data collection between 2004 and 2008, and another set of 5 annual rounds is being implemented through 2013.

One of the additional benefits of conducting a continuous DHS, compared to a conventional DHS, is an opportunity to build capacity in the survey implementing agency which is typically the national statistical office in most countries. By establishing permanent survey teams in the implementing agency, the processes of implementing continuous DHS may improve capacity for survey design, field implementation, data processing, data analysis, and dissemination of results. In particular, the continuous DHS also involves data collection over 9-10 months followed by retraining of interviewers each year. Through continuing training and supervision of a smaller number of interviewers, compared to that in the conventional DHS, quality of data may also improve. An interim evaluation of the Peru CDHS reported strong team spirit among interviewers, but impact on data quality is yet inconclusive based on data from the first 3 rounds (Becker and Pullum, 2007). In addition, PDAs were introduced in 2008 as means of collecting data instead of (and in some cases, in addition to) the paper questionnaires. Although PDAs produce cleaner data, the actual effect on quality has not been assessed. Nevertheless, the popularity of this approach to data collection seems to be soaring. Under the DHS project, PDAs were used only in Peru and Armenia in 2008, but 7 countries are in the pipeline to use PDAs or tablets in 2010.

The purpose of this study is to compare data quality indicators between four conventional DHS (1986, 1991-92, 1996, and 2000) and five CDHS (2004-2008) conducted in Peru. Selected indicators of data quality will be measured to assess age displacement of women and children and digit preference in reporting selected continuous variables. Also, given increasing popularity

of the PDA use, this paper will pays some attention to the data quality effects of the introduction of PDA in 2008.

Data

Demographic and Health Surveys

The DHS are nationally representative cross-sectional household surveys on population, health, and nutrition. The surveys collect background characteristics of households and all members of the households such as age, sex, education attainment. Then, all ever-married women 15 to 49 years of age in the households – hereafter referred to as eligible women – are interviewed to collect detailed information on health and population. The surveys also collect full birth histories to provide data for childhood mortality. All eligible women are asked to report all children ever born and full birth history for each live-born child. The birth history data include information about the date of birth, survival status, and age at death, if dead, of each child. Age at death is reported in days for deaths in the first 28 days of life, in months for deaths between 1 and 23 months, and in years for deaths at ages 2 and over.

In addition, among eligible women who had live birth in the last 5 years before the survey, detailed information on maternal and child health is collected for all pregnancies during the period and all living children under age 5 years. In Peru, four conventional DHS were conducted in 1986, 1991-92, 1996, and 2000.

Continuous Demographic and Health Surveys

The Peru CDHS included 5 annual rounds between 2003 and 2008. The survey used a questionnaire which was roughly similar to that used in the 2000 Peru DHS. During each round, the survey was conducted in one-fifth of the sample clusters selected for the 2000 Peru DHS and interviews were conducted in about 6000 households each year. The survey was initially conducted by seven survey teams (each consisting of a team supervisor, two interviewers and in later rounds a biomarker specialist), supervised by 3 national supervisors, who were full-time staff at the national statistical office. The survey teams were stationed permanently throughout the country and were brought together annually for training/retraining.

Analysis Plan

Measurement of data quality: age displacement by interviewers

We will examine indicators for two common data quality problems: age displacement of women and children by interviewers, and digit preference by interviewees in reporting age at child death. Age displacement of women and children across eligibility boundaries (i.e., transfer of eligible ages to 14 years or younger or 50 years or older among women; and transfer of 0-4 years to 5 years or older among children) reduces the number of eligible women and children for detailed data collection and thus substantially decreases workload for interviewers. In order to measure age displacement of women, we will estimate the proportion of ages 15-19 displaced downward and the proportion of ages 45-49 displaced upward, using a method which examines distributions of four 5-year age groups across the eligibility boundaries (i.e., 5-9, 10-14, 15-19, and 20-24 for downward transfer; and 40-44, 45-49, 50-54, and 55-59 for upward transfer) with 2 simple assumptions (Pullum 2006; Johnson 2009).

Age displacement of children (i.e. upward transfer of 0-4 years to 5 years or older) can be caused by interviewers to reduce work load but also by digit preference among respondents. Thus, we will calculate and compare heaping indices at 5, 10, and 15 years of age (Goldman 1985).

Measurement of data quality: digit preference in reporting

In order to assess levels of digit preference on specific age at death (i.e., heaping), we will examine the distribution of deaths around day 7 and month 12 – two ages with most prevalent digit preference as well as most problematic consequences of heaping leading to underestimation of early neonatal deaths and infant deaths, respectively. A heaping index at each of the two ages will be calculated, using deaths during 0-4 and 5-9 years before each survey (Hill 2006; Curtis 1995).

Comparison of data quality

The main research question of this study is whether data quality differs between conventional and continuous DHS. Age displacement of women and children reflects poor performance by interviewers, whereas digit preference in reporting age at death implies interviewees' reporting errors as well as lack of probing by interviewers. All data quality indicators will be calculated across 4 conventional (1986, 1991-92, 1996, and 2000) and 5 continuous surveys (2004-2008). Differential data quality between conventional and continuous DHS will be assessed, using descriptive estimation as well as graphical analyses.

In addition, for digit preference in reporting child age at death, previous studies reported higher heaping associated with longer recall intervals (Curtis 1995). We will further assess any difference between conventional and continuous surveys vary by recall period: indices during 0-4 vs. 5-9 years before the survey.

We will also compare the data quality indicators across 5 continuous survey in order to assess any effects of the introduction of PDA in 2008 on data quality.

Study Implications

There has been increasing demand for annual estimates of population and health indicators as well as capacity building to conduct quality surveys from donors as well as host countries. Findings from this study will provide knowledge in data quality in Peru CDHS in comparison to conventional DHS, and implications can be expanded to other settings where the continuous survey methodology is being considered. Further, the data quality effects of the introduction of PDA in 2008 will be also discussed.

References

Becker S, Pullum T. 2007. External evaluation of the Peru continuous survey experiment (Report No. 07-001-48). Washington, D.C. Global Health Technical Assistance Project.

Curtis S. 1995. Assessment of the quality of data used for direct estimation of infant and child mortality in DHS-II surveys. 3. Calverton, MD, Macro International Inc. DHS Occational Papers.

Goldman N, Rutstein S, and Singh S. 1985. Assessment of the quality of data in 41 WFS surveys: A comparative approach. WFS Comparative Studies, No. 44. Voorburg, Netherlands: International Statistical Institute.

Johnson K, Grant M, Khan S, Moore Z, Armstrong A, Sa Z. 2009. Fieldwork related factors and data quality in the Demographic and Health Surveys Program. DHS Analytical Studies No. 19. Calverton, Maryland: Macro International Inc.

Hill K, Choi Y. 2006. Neonatal Mortality in the Developing World. Demographic Research. 2006;14:429-452.

Pullum T. 2006. An Assessment of Age and Date Reporting in the DHS Surveys, 1985-2003. Methodological Reports No. 5. Calverton, Maryland: Macro International Inc.