



USAID
FROM THE AMERICAN PEOPLE

The Causes and Consequences of Errors in Demographic Data

March 20, 2012

Thomas Pullum

Demographic and Health Surveys

tpullum@icfi.com

Outline

- Introduction
- Causes of demographic errors
- Methods for detecting errors
- Consequences of errors

Comments are welcome. Goal is to provide a context for the other sessions

Introduction

- The Demographic and Health Surveys (DHS)—description and history
- DHS is a project of USAID (U.S. Agency for International Development), contracted to ICF International. Current five-year funding cycle will end on September 30, 2013.
- My own background—academia until Dec. 2009; with USAID Jan. 2010 through May 2011; joined DHS June 2011

Introduction: domain of “demographic data”

- We are talking entirely, or almost entirely, about information that is objective rather than subjective: dates and counts. These are things that a continuous and complete registration system could record with almost complete accuracy. Sessions today will be on mortality, nuptiality, and fertility.
- DHS and MICS surveys include some information about knowledge, attitudes, and preferences that are inherently subjective and we will not talk much about them today.

Introduction: “demographic errors”

- An error would be an estimate of a population value--a level or a differential or a change--that is outside the range of statistical variability around the population value.
- We do not know the population value, so we usually do not know whether there is a real error, but data quality assessments look at consistency and plausibility to make a judgement.

Causes of demographic errors (incomplete)

- Sampling frame is not valid
- High level of non-response or missing data, and it is non-random, resulting in bias
- The interviews are retrospective, with potential for (1) selective recall; (2) differential survival of respondents; (3) in the case of a marriage history, the revision of an earlier status
- The questions are not clear or not translated correctly
- The interviewers are not adequately trained, supervised, or paid
- The respondent really doesn't know

Methods for detecting errors

- Various internal and external consistency checks
- Demographic modeling, using age and cohort information
- Check for non-random patterns of nonresponse and missing
- Simple symptoms such as heaping can suggest more serious problems

Relatively recent innovations

- *Statistical models.* These models can be multivariate and can produce standard errors that are adjusted for the sample design. Major improvement over non-statistical alternatives.
- *Automatic field editing.* Increasing use of electronic devices for data entry and editing in the field. Mixed blessing from data quality perspective, because the simple symptoms are largely removed even if the fieldwork is sloppy. However, adjustments are tracked and interviewers with many automatic corrections may be re-trained or fired.

Consequences of errors

- Who are the users or clients?
- Why are the data collected?
- What are the costs of errors or misinterpretations?
- Efforts to reduce errors and misinterpretations

Who are the users or clients?

- DHS is a project of the U.S. Agency for International Development, USAID
- Within USAID there are two important funding sources: USAID/Washington , specifically the Bureau for Global Health, and the USAID Missions
- Other donors and agencies are involved, either directly or indirectly, such as UNICEF, WHO, World Bank, UNAIDS

USAID: the Bureau for Global Health

- The Bureau for Global Health has three offices, all of which contribute (literally, in \$\$, to the project):
 - (1) The Office of Population and Reproductive Health, PRH. Formerly called just the Office of Population. The World Fertility Survey began in 1973 and DHS began in 1985 in this office, in order to track fertility preferences, fertility, and use of family planning.

USAID: the Bureau for Global Health

(2) The Office of HIV/AIDS, or OHA, usually just called HIV. This office has grown substantially over the years. It has had an even larger role since the creation of PEPFAR (President's Emergency Plan for AIDS Relief) about ten years ago. It includes interests in knowledge of HIV/AIDS, risky behavior, and seroprevalence. It probably has the most influence on the choice of countries.

USAID: the Bureau for Global Health

(3) The Office of Health, Infectious Diseases, and Nutrition, or HIDN. Its interests include under-five mortality, maternal mortality, ante-natal care, immunizations, and nutrition (hence the height and weight measurements and recent consumption of foods and liquids). Currently under-funded, despite its importance for the questionnaire.

“Cross-cutting themes” at USAID, such as a “gender perspective”, bring in topics such as domestic violence.

USAID Missions

Increasingly, USAID is decentralized, with more decisions and funding coming from the Missions (country offices). The Missions represent the interests of the countries, usually the Ministry of Health, ministries of development, women and children, etc., as well as the interests and experience of the Mission staff, both the Americans and their in-country counterparts.

Why are the data collected?

- The full title of the project is MEASURE DHS.
- MEASURE or measure is a long acronym; the first two letters stand for “monitoring” and “evaluation”
- “monitoring” and “evaluation” are also known as “M&E”, a field that is mostly about constructing indicators of program performance and program impact

Why are the data collected?

- You may not have realized it, but the main purposes of the DHS and MICS surveys are to monitor and evaluate interventions, as well as to help in the targeting and design of interventions.
- You may not be using the data for that purpose, but without such use, primarily by the staff of DHS and UNICEF themselves, the data would not be collected.

What are the costs of errors?

- The DHS project is driven by the use of the data to construct estimates of levels and changes in a wide range of indicators.
- Errors can be real or perceived. This is an important distinction.
- First possibility: Real error, an error in the direction that falsely indicates program success. **False positive.**
- Result can be unjustified continuation of a strategy that IS NOT working

What are the costs of errors?

- Second possibility: Real error, an error in the direction that falsely indicates program failure. **False negative.**
- Result can be unjustified continuation of a strategy that IS working

Many examples. Two successive surveys in the Philippines gave a false impression that the TFR had not declined, despite a surge in the FP program. Was used by opponents to undermine the FP program. The problem was that the second survey had an old sampling frame.

Efforts to reduce errors

- DHS periodically has data assessments. There will be two in 2012:
 - (1) An integrated analysis of errors in fertility and child mortality.
 - (2) An assessment of the quality of the adult and maternal mortality estimates.
- DHS also is trying to automate the field editing, as described above.
- Every survey includes a set of data quality tables.

What are misinterpretations?

- Misinterpretations arise because statistical variation is not perceived as such.
- USAID staff in Washington have a good understanding of statistical variability.
- Pressure to push the data too far and potentially misinterpret the estimates come from Mission staff and country counterparts and from NGOs.

What are the costs of misinterpretations?

- Misinterpretations are potentially as damaging as REAL errors.
 - (1) An estimate can be misinterpreted as indicating program success but is a false positive.
 - (2) OR an estimate can be misinterpreted as indicating program failure but is a false negative.
- The costs of these misinterpretations are similar to the costs of real errors.

Efforts to reduce misinterpretations

- Standard procedures:
- TFR for past three years
- Under-five mortality for past five years
- Maternal mortality for past seven years
- Rates can be given for only the FIRST sub-national unit (“region”), and only if region (and urban/rural) are the sampling domains.

Efforts to reduce misinterpretations

- Increasingly, the country reports include statistical tests and confidence intervals.
- Each survey includes a national seminar at which results are presented and discussion is possible.
- Special briefings and other meetings for journalists.
- Attempts to develop in-country expertise with workshops for various users at universities, national statistical offices, ministries, and USAID Missions.

Fin

As I said, our concern today is mainly with methods to detect and adjust for errors in the data, but we should be aware of the context of this work.

I will try to summarize/synthesize at the end of the day.

Thank you!