(Extended Abstract)

Differential reporting errors in birth histories by socioeconomic group

Yoonjoung Choi and May Sudhinaraset
Johns Hopkins University

Background

An estimated 10 million children under 5 years of age die each year globally, primarily in developing countries,[1] and the Millennium Development Goals aim to reduce under-five mortality by two thirds from 1990 to 2015.[2] Few countries with limited resources, however, have complete vital registration systems which can provide accurate estimates of under-five mortality for monitoring the progress. Standardized birth history data collected in large-scale household surveys – including the World Fertility Surveys and, more recently, the Demographic and Health Surveys (DHS) – have provided information on childhood mortality in many developing countries.[3] Further, with increasing attention to poverty and health inequalities,[4,5] detailed household background information collected in such surveys has been a critical resource to study inequalities in child health and under-five mortality.[6-10]

Nevertheless, although birth history data have been invaluable to improve our knowledge of overall levels and patterns of childhood mortality in many developing countries, data quality issues in birth histories need to be examined further. Two major reporting errors undermine data quality: omission of births of those who died very early during the neonatal period, and digit preference in reporting age at deaths (i.e., heaping) such as on day 7 and month 12 of age. While these errors have been examined mostly at a national level[11] and have been monitored routinely in DHS,[12,13] few studied differential reporting errors across sub-national socioeconomic groups. Any differential level of reporting errors across such groups will bias measurements of inequalities in age-specific mortality as well as differential age patterns of childhood mortality.

The primary purpose of this study is to examine differential reporting errors in birth history data by socioeconomic sub-group, using selected 34 DHS conducted in 21 countries. Specific aims include: (1) to calculate selected data quality indicators by household wealth status and maternal education, and (2) to assess patterns of data quality across socioeconomic status.

Data

Demographic and Health Surveys

The DHS, nationally representative cross-sectional household surveys, have colleted information on the population, health, and nutrition of women and children in over 60 countries since the late 1980s. The surveys collect full birth histories from 15 to 49 year old women in sampled households, providing data for direct estimation of childhood mortality. The respondent is 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. These data provide a basis for direct estimation of child mortality, specific by age and time period. Data also include basic demographic and socioeconomic information about each of all 15 to 49 year old women in sampled households and her household, including their education attainment, household ownership of selected material assets, housing condition, and whether the household is located in a rural or urban area.

Analysis data

We systematically reviewed 165 DHS available as of May 2009, and excluded surveys conducted during the first phase in the late 1980s due to potentially substantial improvement in survey implementation and data quality.[14] We purposely selected 34 surveys with sufficient number of neonatal and infant deaths in order to examine data quality at sub-national levels (Appendix 1). The mean under-five mortality rate was 169 per 1000 live births (range [58, 330]), and 18 surveys were conducted in sub-Saharan Africa.

Analysis Plan

Measurement of data quality

We examined four data quality indicators to measure degree of heaping and omission of births. In order to assess levels of digit preference on specific age at death, we examined 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. We calculated a simple heaping index at each age. With an assumption that the number of deaths by day/month is a linear function in the absence of digit preference, heaping indices were calculated as below:

$$\text{HI(d7)} = \frac{D(d7)}{\{\text{D}(d5)+\text{D}(d6)+\text{D}(d7)+\text{D}(d8)+\text{D}(d9)\}/5}\text{; and} \quad (1)$$

$$\text{HI(m12)} = \frac{D(m12)}{\{\text{D}(m10)+\text{D}(m11)+\text{D}(m12)+\text{D}(m13)+\text{D}(m14)\}/5} \quad (2)$$

, where HI refers to heaping index, D is the number of deaths reported at exact age in parenthesis, d refers to day, and m refers to month.

In order to study potential omission of live births, we examined: (1) the ratio of Early Neonatal Mortality Rate (ENMR) to Neonatal Mortality Rate (NMR), compared to overall level of NMR, and (2) the ratio of NMR to Infant Mortality Rate (IMR), compared to overall level of IMR. Quantifying magnitude of omission of live births is not straightforward. However, substantial omission may be suggested from implausibly low ratios of mortality rate (ENMR/NMR and NMR/IMR) relative to overall mortality rate (NMR and IMR, respectively), compared to the associations observed in historic data from developed countries with presumably good data quality.[11] We obtained linear regression fitted values of each of the ratios (ENMR/NMR and NMR/IMR) as a function of the mortality rate (NMR and IMR, respectively) using historic data from England and Wales.[11] Any observed DHS ratio which is substantially lower than the fitted values indicates omission of live births relative to any omission reflected in the historic data, even though we do not know the absolute magnitude of omission of live births in the historic England and Wales data.

In this paper, mortality rates refer to probabilities of dying between two exact ages.[i] as follows. A cohort probability approach was used to calculated ENMR and NMR – the number of early neonatal deaths and neonatal deaths, respectively, divided by the number of live births during a defined period. Any child who was less than 7 and 28 day old at the time of survey was excluded from the calculation of ENMR and NMR, respectively. On the other hand, we calculated IMR, using a synthetic cohort method,[15,16] in order to reflect the mortality risk which substantially varies with age at exposure during the first year of life.

Measurement of socioeconomic characteristics

We measured household wealth and maternal educational attainment as indicators of socioeconomic characteristics. To measure household wealth, we either calculated a household wealth index score, using principal component analysis of household assets and housing conditions,[17] or, when available, used the score variable created by DHS in recent surveys. Households were ranked based on the index score and categorized into quintiles. We further categorized households into quartiles and tertiles for sensitivity analyses. Maternal education attainment was categorized into two: < vs. ≥ completion of primary education, for which the exact years slightly vary by country.

Assessment of differential pattern of data quality

Data quality indicators will be calculated across household wealth and maternal education sub-groups for two consecutive 10-year periods before each survey (i.e., 0-9 and 10-19 years before the survey), based on the difference between the date of survey and the date of death. Analysis will be restricted to group-specific observations with at least 30 neonatal deaths in each reference period, since a low number of deaths was reported in high socioeconomic subgroups (i.e., the highest wealth quintile) in some surveys. Differential levels of data quality indicators will be assessed across household wealth and maternal education subgroups, using descriptive estimation as well as graphical analyses. Further, differential socioeconomic patterns of data quality will be explored by region as well as the level of overall under-five mortality.

Strengths and limitations

Our study will examine differential levels of reporting errors in birth histories by socioeconomic group, using selected 34 DHS. Analysis data will include diverse populations in terms of region and overall levels of childhood mortality. Investigating patterns of reporting errors at the sub-national level using such data will provide valuable

[i] Definitions of age-specific mortality rates are as follows. ENMR: probability of dying between 0-6 days of age; NMR: probability of dying between 0-27 days of age; and IMR: probability of dying between 0-11 months of age.

knowledge to understand further measurements of sub-national variation in levels of age-specific childhood mortality and age patterns of childhood mortality in developing countries.

However, in addition to reporting errors, the built-in maternal survival bias in birth histories (i.e., birth history information is collected only from surviving mothers) may also contribute biased measurement of inequalities in childhood mortality. The selection bias underestimates overall childhood mortality,[18,19] and levels of this selection bias may be reversely associated with socioeconomic characteristics due to socioeconomic gradient in adult female mortality, resulting in underestimation of the inequality in overall under-five mortality. Further, if the degree of correlation between maternal and child survival varies by age of child, the selection bias will also affect measurements of age-patterns of mortality. Nevertheless, in populations without HIV/AIDS epidemic, female mortality between 15-49 years of age is typically low and overall levels of under-five mortality may not be underestimated substantially.

Study Implications

Increasing attention has been given to poverty and inequalities in child health. Our study will address differential data quality in birth histories across socioeconomic groups. Findings will provide knowledge in measurements of inequalities in levels of age-specific childhood mortality as well as variation in age-patterns of childhood mortality, utilizing birth history data such as DHS. Such contribution will be valuable to better monitor progresses in decreasing the inequality in health outcomes in developing countries.

Reference List

(1) Black RE, Morris SS, Bryce J. Where and why are 10 million children dying every year? *Lancet.* 2003;361:2226-2234.

(2) United Nations. General Assembly, 56th session. Road map towards the implementation of the United Nations millennium declaration: report of the Secretary-General. UN document no. A/56/326. 2001. New York, NY, United Nations.

(3) Boerma JT, Sommerfelt AE. Demographic and health surveys (DHS): contributions and limitations. *World Health Stat Q.* 1993;46:222-226.

(4) Wagstaff A, Bustreo F, Bryce J, Claeson M. Child health: reaching the poor. *Am J Public Health.* 2004;94:726-736.

(5) Wagstaff A. Poverty and health sector inequalities. *Bull World Health Organ.* 2002;80:97-105.

(6) Wirth ME, Balk D, Delamonica E, Storeygard A, Sacks E, Minujin A. Setting the stage for equity-sensitive monitoring of the maternal and child health Millennium Development Goals. *Bull World Health Organ.* 2006;84:519-527.

(7) Gwatkin DR, Rutstein S, Johnson K, Suliman E, Wagstaff A, Amouzou A. Socio-economic differences in health, nutrition, and population within developing countries: an overview. *Niger J Clin Pract.* 2007;10:272-282.

(8) Wagstaff A, Van DE. Overall versus socioeconomic health inequality: a measurement framework and two empirical illustrations. *Health Econ.* 2004;13:297-301.

(9) Wagstaff A. Socioeconomic inequalities in child mortality: comparisons across nine developing countries. *Bull World Health Organ.* 2000;78:19-29.

(10) Gwatkin DR. Health inequalities and the health of the poor: what do we know? What can we do? *Bull World Health Organ.* 2000;78:3-18.

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

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

(13) Arnold, F. Assessment of the quality of birth history data in the Demographic and Health Surveys. In *An assessment of DHS-I data quality*. 1. 1990. Columbia, MD, Institute for Resource Development/ Macro Systems, Inc. DHS Methodological Reports.

(14) Marckwardt, A. M. and Rutstein, S. Accuracy of DHS-II Demographic data: gains and losses in comparison with earlier surveys. 19. 1996. Calverton, MD, Macro International, Inc. DHS Working Papers.

(15) Somoza, J. L. Illustrative analysis: Infant and child mortality in Colombia. 38. 1980. Voorberg, Netherlands, International Statistical Institute. WFS Scientific Reports.

(16) Rutstein, S. Infant and child mortality: levels. trends, and demographic differntials. 43. 1984. Voorburg, Netherlands, International Statistical Institute. WFS Comparative Studies.

(17) Filmer D, Pritchett LH. Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India. *Demography.* 2001;38:115-132.

(18) Nakiyingi JS, Bracher M, Whitworth JA et al. Child survival in relation to mother's HIV infection and survival: evidence from a Ugandan cohort study. *AIDS.* 2003;17:1827-1834.

(19) Zaba B, Whitworth J, Marston M et al. HIV and mortality of mothers and children: evidence from cohort studies in Uganda, Tanzania, and Malawi. *Epidemiology.* 2005;16:275-280.

Appendix 1. Selected 34 study surveys, number of childhood deaths, and under-five mortality rate*

Country	Survey year	Births	Under-five deaths	Infant deaths	Neonatal deaths	Early neonatal deaths	Under-five mortality rate (per 1000 live births)
Bangladesh	1994	6968	714	577	350	224	218
Bangladesh	2000	6744	501	430	280	201	187
Benin	2006	15918	1388	998	493	394	151
Cambodia	2000	8748	964	830	340	243	133
Coted Ivoire	1994	6755	784	558	281	212	150
Egypt	1995	12007	858	750	359	226	138
Ethiopia	2000	10763	1309	952	490	352	211
Ethiopia	2005	9749	854	663	335	232	199
Guatemala	1995	9840	590	498	273	208	68
Guinea	1999	5750	790	553	278	208	265
India	1993	60034	4939	4143	2609	1809	144
India	1999	56259	4240	3502	2248	1654	107
India	2006	51172	2863	2495	1702	1318	95
Indonesia	1991	15568	1175	964	456	290	116
Indonesia	1994	18026	1195	977	495	348	110
Indonesia	1997	17296	896	778	392	285	58
Indonesia	2003	16041	699	601	330	255	63
Indonesia	2007	18456	749	638	371	283	89
Malawi	2000	11807	1555	1102	465	328	247
Malawi	2004	10801	1055	805	318	233	187
Mali	1996	10158	1730	1139	587	415	315
Mali	2001	12932	1981	1332	694	480	330
Mali	2006	14072	1795	1208	623	489	242
Mozambique	1997	7005	932	715	310	231	201
Mozambique	2003	10238	1193	939	360	227	254
Nepal	1996	7292	693	553	353	223	228
Nigeria	2003	5972	843	559	292	214	234
Pakistan	2007	9064	725	649	441	338	92
Peru	1996	17405	944	790	447	298	78
Rwanda	2000	7851	1065	761	322	240	219
Rwanda	2005	8543	891	661	311	222	172
Senegal	2005	10842	839	609	357	275	150
Tanzania	2004	8480	709	524	260	201	160
Togo	1998	6985	723	508	258	208	141

*during the 5-year period preceding the survey, calculated by the authors.