Siblings Count: The Extent and Effect of Double Counting in the Sibling Survival Method using the Demographic and Health Surveys

Megan Costa, University of Pennsylvania

Abstract

Many low-resource settings lack accurate vital registration data, resulting in increased use of cross-sectional surveys to estimate all-cause mortality. Recent developments in estimating adult mortality from survey data include improvements to the sibling survival method. Large families of surveys employ sampling strategies that interview all present adults in the household. While cost effective, this strategy introduces potential duplication of sibling records within the household.

Using DHS data, a hierarchical algorithm was developed to identify multiple sibling respondents within the household. The number of sibling respondents per sibling group was quantified and proportion of excess respondents (non-unique by sibling group) calculated. Mortality rates with and without excess sibling records were calculated using survival time analysis.

Preliminary results using seven DHS reveal heterogeneous sibling record duplication across countries. Excess respondents ranged from 3.9% in Burkina Faso to 9.5% in Haiti, giving rise to 4.4% and 10.6% of sibling observations, respectively. An analysis of the 2007 Zambia DHS suggests varying effects of record duplication on mortality rates by age.

This manuscript predicts the proportion of duplicate respondents and sibling observations will vary by survey to reflect family structure and living arrangements within each country. After excluding duplicate records, five-year mortality rates from ages 15-35 are predicted to vary by country and context, and confidence intervals will widen to reflect the true number of observations. Using this assessment of sibling record duplication in the DHS, researchers using the sibling history method are recommended to quantify and control for duplicate records and incorporate the increased uncertainty in their estimates.

Introduction

Globally, all-cause mortality has been declining in children and adults for over 50 years (Rajaratnam 2010; UNPD 2010; Lozano 2011). With increased resources allocated to global health and mortality reduction, monitoring baseline mortality levels and change is critical to ensuring efficient use of development assistance for health and monitoring progress toward the Millennium Development Goals (Murray 2011). Despite the necessity to monitor all-cause mortality, only 26% of the world population lived in a country with complete death registration from 1995-2004 (Mahapatra 2007). China, India, and most African WHO member states had <50% vital registration completeness from 1995-2004. Some efforts have been made to improve the quality of vital registration globally, however, data improvements have been limited in scope (Mahapatra 2007). Health researchers and policy makers must therefore make use of incomplete and low-quality data supplemented by population health surveys to generate estimates of all-cause mortality. Methods to analyze survey data have improved over the last several decades, with recent developments in adult mortality estimation bringing about increased utilization of sibling histories.

The sibling history method used to estimate adult mortality is analogous to the direct method to estimate child mortality from complete birth histories. During a survey, respondents complete a module in which they report the sex, survival status, and age of each sibling related biologically through the same mother. For siblings who have died, age at death is reported. Data from sibling histories are only valid when information on the entire sibling group has been collected. Sibling exposure time and deaths are tabulated by five-year age group and period and death rates are calculated. This method can be used to measure mortality for an age range similar to respondent ages, as siblings will be on average the same age (Timaeus 2012). Adult mortality estimates from sibling histories do not assume a closed population to migration and can generate mortality estimates for more than one period, as opposed to using deaths in the past year as commonly collected in censuses (Timaeus 2012).

Several methodological issues have been identified and must be accounted for when using the sibling survival method to generate estimates of adult mortality. Trussell and Rodriguez (1990) identified three central limitations to the sibling survival method; first, high-mortality sibling groups are under-represented because sibling groups without survivors will have no respondent in the survey. Conversely, sibling groups with low mortality will be over-represented because multiple siblings can be eligible respondents, thus double-counting the mortality experience of that low-mortality sibling group. Lastly, the respondent is not counted as part of the denominator and without their survival status counted, mortality estimates are artificially inflated.

Mathematically, these three limitations cancel out, as long as mortality and sibling group size are not correlated and groups of siblings have the same underlying probability of death (Trussell and Rodríguez 1990; Reniers 2010; Timaeus 2012). Though previous studies posit a correlation between sibling group size and mortality and introduce an adjustment for this relationship, a recent assessment concludes that when restricting the sample to adult siblings and adult deaths only, the correlation between sibling group size and mortality disappears (Gakidou and King 2006; Masquelier 2011).

When respondents report on siblings of the opposite sex, weighting is required to avoid bias (Timaeus 2012). Respondents reporting on deaths of the opposite sex experience a differential risk of death. Bias may be introduced "if the mortality of siblings of one sex is associated with the number of siblings of the opposite sex that report on them" (Timaeus 2012). So for women reporting on brothers, each report must be weighted by the inverse of her number of living sisters, which basically assumes that mortality of individuals in sibling groups with no respondents alive to report is the same as mortality of other sibling groups in the population (Timaeus 2012).

In addition to these methodological issues, estimates from the sibling history method can be implausibly low due to omission of sibling deaths as the respondent recalls events farther back in time from the survey date (Timæus and Jasseh 2004). Other applications of this method have resulted in underestimates (Gakidou, Hogan et al. 2004). For this reason, the sibling history method has been referred to as a "lower bounds" of adult mortality (Masquelier 2011). To avoid artificially low mortality rates further back in time, reference periods should be limited to a maximum of 10 years in the past. Often a maximum of 7 is used, broken up into 3- and 4-year periods prior to the survey. Additional bias can be avoided if mortality rates are calculated for a limited set of ages, usually from a minimum of age 15 to adulthood (Timaeus 2012).

With these methodological developments, the sibling survival method has been applied to estimate adult mortality in a range of settings. The sisterhood method has been used to provide empirical data for estimating maternal mortality and assessing progress toward Millennium Development Goal 5, a target set by the World Health Organization to reduce by 75% the maternal mortality ratio by 2015 (Hogan, Foreman et al. 2010; Lozano 2011). The sibling survival method has also been used to assess adult mortality post-conflict; a study quantifying excess mortality during the 1994 genocide in Rwanda found a higher proportion of excess deaths among urban and educated males (de Walque 2010). An analysis of the World Health Surveys (WHS) undertaken in 2008 analyzed estimated war deaths over the course of 50 years in 13 countries, and recently the sibling history method was used to estimate war deaths in Iraq by the Iraq Family Health Survey Group (IFHS 2008; Obermeyer, Murray et al. 2008). Many of these studies employ slightly different versions of the sibling survival method as refinements were introduced over time, but recent years have seen increased use of the method to generate adult mortality estimates critical to understanding past and present trends and in mortality. Given the increased use of the sibling history method, it is critical for researchers and consumers of estimates to have a clear understanding of what biases, if any, are inherent in the method and how they can be ameliorated.

The Problem

Demographic and Health surveys (DHS) were established in 1984 by the US Agency for International Development (USAID) to gather standardized data on fertility and household-level indicators in data-sparse settings. Information on all residents in the household is gathered from the head of household, and eligible women are selected from this roster. Using this method, it is possible to have multiple female respondents in each household. The DHS include sibling history modules in over 50 African surveys, however, the respondents are usually limited to women of reproductive age (15-49) and the total number of respondents and corresponding sibling observations are too few to generate age and period-specific death rates that do not require smoothing (Hill, Choi et al. 2005; Reniers 2010). Sibling history modules were originally included in the DHS to facilitate the sisterhood method to estimate maternal mortality. In the sisterhood method, women were asked about the survival status of their sisters and if any deaths were related to maternal causes. The DHS now include questions on siblings of both sexes. Sibling history questions are asked of male respondents in only 20% of DHS, but can be analyzed using the same method as female respondents (Reniers 2010, Timaeus 2012).

The DHS is one of the largest survey families analyzed using the sibling survival method. The sampling strategy of the DHS is to interview all eligible respondents in a household; however, no indication is made as to whether these respondents are actually siblings. In previous works, when the sibling history method is applied to DHS no attempt is made to quantify the duplication of sibling groups from sibling respondents within the same household (Masquelier 2011). In an analysis of the 2000 Rwandan DHS, de Walque and colleagues avoid this issue by randomly retaining only one respondent per household per total sibling group size (de Walque 2010). In this study, a hierarchical algorithm is developed to categorize respondents as unique or duplicate observations and will be applied to DHS to quantify duplication of sibling history records. Adult mortality from ages 15 to 35 (${}_{20}M_{15}$) and their confidence intervals for the seven years prior to the survey will be calculated with and without duplicate records to determine the extent and effect of duplicate records on mortality rates. Though the effect of duplication should theoretically cancel out, this study seeks to determine whether duplication can and should be corrected for when generating adult mortality estimates using this method.

Data and Research Methods

A hierarchical algorithm for grouping siblings utilized household-level indicators collected in the household roster and individual questionnaire to match individual respondents as siblings. This process used two standard DHS files. The household member list, which contains one record per household member and multiple records per household, was linked with the individual questionnaire, which contains one record per respondent and all of their sibling history data. This process is limited to matching siblings within household due to available data.

Respondents with no eligible siblings of the same sex were identified using the individual questionnaire. Single eligible respondents had no risk of respondent-siblings, and were excluded after categorization. Respondents not categorized as single eligible respondents in the previous step proceeded to the next, where mothers' line number was used to identify siblings. Mothers' line number as collected in the household roster was merged onto the individual respondents' records. Individuals in the same household sharing the same mothers' line number were categorized as siblings. Due to the potential for inaccurate classification as single eligible respondents during this step because of a high degree of missing values, those identified as single respondents in addition to those yet unclassified proceeded to the next step.

Following the application of mothers' line number, relationship to the head of household was used to group uncategorized siblings. The household roster was merged onto the individual questionnaire, generating an indicator for the respondents' relationship to the head of household. Respondents categorized as sons or daughters of the head of household were grouped as siblings within the household. Next, respondents categorized as brothers or sisters of the head of household were grouped as siblings, including the household head themselves.

After classification of respondents as single eligible respondents or sibling groups using mothers' line number or relationship to head of household, individual characteristics including age group and number of siblings alive and dead were used to match respondents within the household. For respondents unclassified by the previous steps, sibling history questions from the individual questionnaire were used to generate a list of the entire sibling group reported by a respondent, including a record for the respondent. A list of the yet-unclassified respondents was merged to this sibling group list using household identifiers, age group, number of eligible siblings of the same sex, and total number of living and dead siblings.

This hierarchical algorithm results in a count of the "linked siblings," or the number of respondents linked to the same mother. Because male and female respondents are analyzed separately in the sibling survival method, "linked siblings" represents the sex-specific number of linked siblings. For each linked sibling group, 1/N of the sibling records will be retained randomly, where N is the total number of linked sibling respondents within the household. Mortality rates will be calculated and compared before and after randomly retaining one observation per sibling group.

Preliminary Results

Table 1 presents the proportion of excess respondents by survey for a pilot application of the algorithm to seven DHS surveys. A respondent is counted as "excess" if they are in excess of the first observation within a household. The proportion of excess respondents varies by country, but does not exceed 9.5% in this sample. Duplicate sibling records are expected to vary by country, as sibling co-residence likely to vary due to differences in family size, socio-economic status, marital patterns, and other country-specific characteristics. A wider application of this algorithm to all DHS with sibling history modules would provide a better understanding of what the true sample size is when applying this method, and how it affects mortality estimates.

Table 1. Proportion of Excess Respondents by Country Year, weighted using DHS weights

Country	Year	Number of Respondents	Excess Respondents N, %
Bolivia	2008	16,939	999, (5.9%)
Burkina Faso	2003	12,477	362, (2.9%)
Cambodia	2010	18,754	1,725, (9.2%)
East Timor	2009	13,137	1,064, (8.1%)
Haiti	2005	10,757	1,022, (9.5%)
Zambia	2007	7,146	329, (4.6%)
Zimbabwe	2005	8,907	428, (4.8%)

Preliminary analysis of mortality by five-year age group for the seven years prior to the survey in Zambia suggests varying effects of duplicate records by age (Table 2). The overall proportion of excess female respondents is low in the Zambia 2007 survey (4.6%), and excluding this proportion results in a 3.72% increase in mortality rate for brothers aged 15-19. Duplication by age of respondent and age of siblings should further be explored to determine whether mortality rates change significantly with and without duplicate sibling records.

Table 2. Zambia 2007 female respondents reporting brothers: mortality rates, ages 15-35

Age Group	Including Excess Respondents		Excluding Excess Respondents		% Difference
	Rate Per 1000	95% CI	Rate Per 1000	95% CI	
15	3.31	2.53, 3.30	3.19	2.41, 4.31	3.72%
20	5.70	4.65, 5.70	5.70	4.62, 7.11	0.07%
25	8.34	7.03, 8.34	8.52	7.16, 10.22	-2.11%
30	18.35	16.09, 18.35	18.28	16.00, 20.98	0.40%
35	25.17	22.95, 25.17	25.14	22.92, 27.64	0.09%

While the overall difference in mortality rates after excluding excess siblings was small, fluctuations age-specific reported mortality rates have an impact on patterns of adult mortality. Mortality rates generated using this method used in national estimates have dramatic socio-political and monetary impacts on countries seeking economic aid. The full version of this manuscript uses additional DHS sites to better quantify the difference in mortality rates, and recommends that hierarchical methods are used to safeguard against inaccurate measures of mortality.

References

de Walque, D., Verwimp, P. (2010). "The Demographic and Socioeconomic Distribution of Excess Mortality during the 1994 Genocide in Rwanda." Journal of African Economies **19**(2): 141-162.

Gakidou, E., M. Hogan, et al. (2004). "Adult mortality: time for a reappraisal." International Journal of Epidemiology **33**(4): 710-717.

Gakidou, E. and G. King (2006). "Death by Survey: Estimating Adult Mortality without Selection Bias from Sibling Survival Data." Demography **43**(3): 569-585.

Hill, K., Y. Choi, et al. (2005). "Unconventional approaches to mortality estimation." DemRes Demographic Research **13**: 281-300.

Hogan, M. C., K. J. Foreman, et al. (2010). "Maternal mortality for 181 countries, 1980-2008: a systematic analysis of progress towards Millennium Development Goal 5." Lancet **375**(9726): 1609-1623.

IFHS, I. F. H. S. G. (2008). "Violence-related Mortality in Iraq from 2002 to 2006." New England Journal of Medicine **358**(5).

Lozano, R., H Wang, KJ Foreman, J.l Rajaratnam, M Naghavi, J Marcus, L Dwyer-Lindgren, K Lofgren, D Phillips, C Atkinson A Lopez, C Murray (2011). "Progress Towards Millennium Development Goals 4 and 5 on Maternal and Child Mortality: an Updated Systematic Analysis"." Lancet **378**(9797): 1139-1165.

Mahapatra, P., K Shibuya, A Lopez, F Coullare, FC Notzon, C Rao, S Szretzer (2007). "Civil Registration Systems and VitalStatistics: Successes and Missed Opportunities." Lancet **370**(9599): 1653-1663.

Masquelier, B. (2011). "Adult Mortality From Sibling Survival Data: Does the Corrected Method Perform Better? ."

Murray, C. A., B Burstein, R Leach-Kemon, K Schneider, M, Tardif, A Zhang, R (2011). "Development Assistance for Health: Trends and Prospects." Lancet **378**(9785).

Obermeyer, Z., C. J. L. Murray, et al. (2008). "Fifty years of violent war deaths from Vietnam to Bosnia: analysis of data from world health survey programme - When one respondent per household reported on siblings' deaths, the estimated total of war deaths in 13 countries from 1955-2002 was 5.4 million, two thirds more than suggested by censuses, routine surveillance, and media reports." BMJ : British medical journal / **88**(7659): 1482.

Rajaratnam, J., Marcus, J. , Levin-Rector, A, Chalupka, A, Wang, H, Dwyer, L, Costa, M, Lopez, A, Murracy, C. (2010). "Worldwide Mortality in Men and Women aged 15-59 Years from 1970 to 2010: a Systematic Analysis." The Lancet **375**(9727): 1704-1720.

Rajaratnam, J., Marcus, J. Flaxman, A, Wang, H, Levin-Rector, A Dwyer, L, Costa, M, Lopez, A, Murracy, C. (2010). "Neonatal, Postneonatal, and Under-5 Mortality for 187 Countries, 1970-2010: a Systematic Analysis of Progress Towards Millennium Development Goal 4." Lancet **375**(9730).

Reniers, G., Masquelier, Bruno, Gerland, Patrick (2010). "Adult Mortality Trends in Africa." Working Paper, OPR: 35.

Timaeus, I. (2012). Estimation of adult mortality from sibling histories. Tools for Demographic Estimation. D. R. H. A. Moultrie TA, Hill KH, Timaeus IM, Zaba B. Paris, International Union for the Scientific Study of Population.

Timæus, I. M. and M. Jasseh (2004). "Adult Mortality in Sub-Saharan Africa: Evidence from Demographic and Health Surveys." Demography **41**(4): 757-772.

Trussell, J. and G. Rodríguez (1990). "A Note on the Sisterhood Estimator of Maternal Mortality." Studies in Family Planning **21**(6): 344-346.

UNPD (2010). World Population Prospects: The 2010 Revision. New York, United Nations.