Bruno Masquelier*

Sibship Sizes and Family Sizes in Survey Data Used to Estimate Mortality

How can mortality be estimated when vital records are inaccurate and incomplete? One widely-applied method is to use sibling survival data to measure adult mortality levels. Bruno Masquelier *assesses the quality of these data by comparing expected sibship size (as deduced from the fertility of the preceding generation) with sibship sizes reported in 109 Demographic and Health Surveys conducted in some 50 countries since the 1990s. He shows that the quality of sibling data is very variable. In a very large majority of surveys, sibship size is underestimated by around 15% on average, but with large differences across surveys. These omissions mainly concern brothers or sisters who died in childhood, so have little impact on adult mortality. Nonetheless, it is important to understand such omissions in order to quantify the biases affecting these indirect mortality estimates.*

Survey data on sibling survival are essential for producing mortality estimates in many countries where the registration of deaths remains incomplete. These data are used widely today to reconstitute trends in both adult (Wang et al., 2012) and maternal mortality (Wilmoth et al., 2012), to estimate conflict-related mortality (Hagopian et al., 2013), or to assess the effects of health programmes (Bendavid et al., 2012).

These data are generally collected via a standardized and quite repetitive questionnaire which begins by asking respondents to give a list of their brothers and sisters born to the same mother, including half-sisters and half-brothers, if any. They are then asked to detail the sex and vital status of each sibling by birth order. Current age is recorded for surviving siblings, and age at death and years since death for those who are deceased. Additional questions serve to identify sisters who died during pregnancy or delivery, or in the two months following childbirth. In other words, these data provide a complete birth history of the respondent's mother, and biographical methods can be used to link

* Centre de recherches en démographie et sociétés, Université catholique de Louvain, Belgium.

Correspondence: Bruno Masquelier, Centre de recherches en démographie et sociétés, Université catholique de Louvain, Place Montesquieu 1, bte L2.08.03, 1348 Louvain-la-Neuve, Belgium, tel. : +32 10 47 26 17, email: bruno.masquelier@uclouvain.be

Population-E, 69 (2), 2014, 221-238 DOI: 10.3917/pope.1402.0221

deaths by age and period to the corresponding periods of exposure. These questions have been included in a number of survey programmes, including more than a hundred Demographic and Health Surveys (DHS) since 1989 (on women aged 15-49), Reproductive Health Surveys by the Centers for Disease Control and Prevention (CDC), and the World Health Survey (WHS) conducted by the World Health Organization (WHO).

Although sibling histories are used increasingly to produce mortality estimates, their quality has rarely been assessed. Methodological studies have focused primarily on modelling of mortality trends (Obermeyer et al., 2010; Timaeus and Jasseh, 2004) and on selection biases resulting from the retrospective nature of the data collected (Gakidou and King, 2006; Masquelier, 2013). Comparatively less attention has been paid to reporting errors such as omitted deaths, inaccurate ages and dates, or problems in identifying maternal deaths (Helleringer et al., 2013). Existing research on this topic has followed four different strands, so to place our study in context, we will start with a summary of their main conclusions.

First, Helleringer et al. (2014) recently assessed the quality of sibling histories by comparing them with data collected longitudinally on a demographic surveillance site (DSS) in southern Senegal. To this end, they organized a DHS-type survey in the DSS, then matched the sibling information collected with the surveillance data at individual level. Focusing on female mortality, they observed that sisters were frequently omitted by respondents. The proportions were 4% for surviving adult sisters, 9% for sisters who died in adulthood and 17% for sisters who had migrated out of the DSS. They also observed a tendency to underestimate age at death, particularly at older ages (45 years and above). These two types of errors tend to cancel each other out, since the omissions introduce a downward bias in mortality levels while age underestimation introduces an upward bias. However, compensation is only partial and measurement errors may go in either direction. Moreover, these results were obtained in a very specific rural area, and further research is needed to determine whether they can be generalized to other countries where adult mortality is higher, for example, and where the structure of the sibships is different.

To detect reporting errors more systematically, several diagnostic exercises were conducted to assess adult mortality rates based on DHS data by comparing them with the estimates published by the United Nations in its World Population Prospects (WPP). All these comparisons suggest that adult mortality based on sibling data is underestimated. It is difficult to draw firm conclusions from such comparisons, however, since the reference mortality rates may themselves be inaccurate. This is especially the case in countries affected by the HIV/AIDS epidemic, for which a complex model of the disease must be used. Intriguingly, sibling data produce mortality levels in Sahelian countries that are very low compared with UN estimates, but the two series are more closely aligned in countries of southern Africa where mortality is dominated by AIDS (Reniers et al., 2011). To date, we do not know if these regional variations correspond to differences in reporting accuracy, or to errors in mortality modelling.

Stanton et al. (2000) adopted a third approach that involved internal consistency tests on survey data from 14 DHS surveys conducted between 1989 and 1995. They examined, for example, how the mean numbers of reported brothers and sisters vary with the respondents' age, and they observed that older respondents tend to report as many or fewer siblings than younger ones. As older women's sibships were formed in a more distant past, at a time when fertility was generally higher, it would appear that these older respondents more frequently omit brothers and sisters, doubtless due to recall problems. They also showed, however, that there are few missing data on the survival of siblings and that ages at the time of the survey or at death are given in more than 95% of cases. Completeness of information does not seem to vary by sex or by years since death, although only reported deaths are concerned. It is still possible that certain deaths in the distant past are simply not reported. In fact, Stanton et al. (2000) detect an abnormally high concentration of deaths in the years running up to the survey, as well as heaping on 5, 10 or 15 years before the survey.

Last, this question of under-reporting of more distant deaths has been examined in another series of studies which compared the mortality levels obtained for the same calendar period but from surveys held several years apart in a given country (Obermeyer et al., 2010; Timaeus and Jasseh, 2004). These comparisons confirm that under-reporting of deaths rises rapidly as the time between the death and the survey date increases. In sub-Saharan Africa, mortality rates for periods earlier than six years before the survey are more than 25% lower than those estimated for the years directly preceding it (Masquelier et al., 2014). The mortality rates can be adjusted accordingly, but only to correct relative under-reporting of deaths, i.e. by comparison with the most recent periods. Estimates remain biased if the omissions also concern deaths that occurred in the years immediately preceding the survey, as was observed in Senegal (Helleringer et al., 2014). In addition, several surveys must be combined in order to implement this comparative approach, so it cannot be used to judge the quality of each survey taken individually.

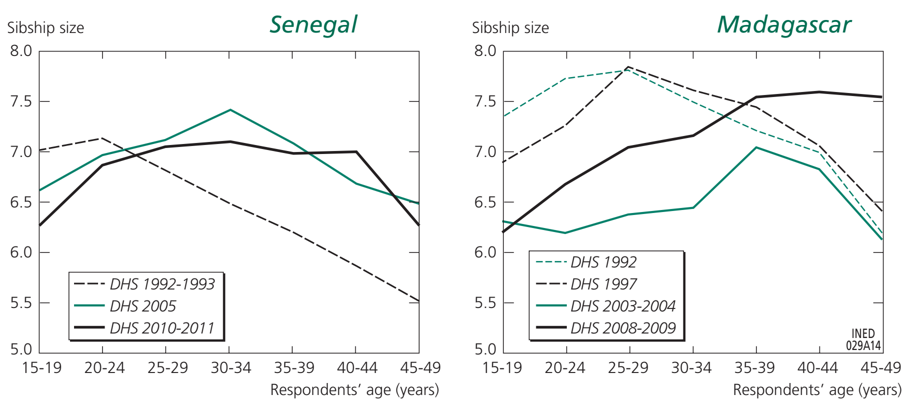
Unfortunately, no gold standards are available to assess the scale of omissions of deceased siblings in each DHS survey. However, the total number of reported siblings (i.e. the sibship size) can be examined in more detail, even though drawing conclusions about mortality estimates is more difficult. As mentioned above, Stanton et al. (2000) look at how this number varies with the respondents' age, and conclude that older respondents omit more siblings than younger ones. But is the number of siblings reported by younger respondents itself plausible? Rather than simply comparing sibship size by age, this article examines sibship size in relation to past fertility. Preston (1976) has shown that an algebraic relation exists between mean number of children and mean sibship size: the mean sibship size is directly dependent on the mean number of children ever born in the previous generation and on the variance of this number. Preston's equivalence (1976) is used here to calculate "expected" sibship sizes. This approach, which complements existing quality diagnostics, provides a means to assess the scale of sibling omissions for each survey taken separately, and for all respondent age groups.

I. Data and method

This analysis makes use of all the standard DHS surveys available in the public domain in December 2013 and which include a module on sibling survival, i.e. a total of 109 surveys of women in 50 different countries. For each five-year age group (from ages 15 to 49), the mean sizes of maternal sibships (brothers and sisters with the same mother), including the respondents themselves, were extracted from these surveys. Only data collected from women are used here, since men were asked about their siblings in just a dozen or so surveys. In any case, there is no indication in these surveys of systematic differences between numbers of siblings reported by men and by women; the correlation coefficient between the two series is 0.94. Merdad et al. (2013) have also shown that adult mortality levels deduced from information reported by men were not significantly different from those deduced from women's reports.

We can begin the analysis with two extreme examples: mean sibship sizes reported in surveys conducted in Madagascar in 1992 and in Senegal in 1992-1993 (Figure 1). We see that in both cases, the sibship size decreases rapidly with the respondent's age. Women aged 45-49 report around 20% fewer siblings than women aged 20-24. This decrease with age is not observed in all surveys,

Figure 1. Reported sibship size by respondents' age, DHS surveys in Senegal and Madagascar



Sources: Demographic and Health Surveys (www.measuredhs.com).

however; the opposite pattern was observed in Madagascar in 2009. Does this mean that the 2009 data are of better quality? Before blaming omissions, two alternative hypotheses should be considered. First, if adult mortality is higher in larger sibships, these large sibships will become increasingly under-represented as respondents advance in age. However, a recent analysis of DHS data has shown that adult survival varies little with the number of adult sisters (Masquelier, 2013), so these small survival differentials, associated with mortality which is itself low at these ages, cannot explain such large variations in the number of reported siblings. This leaves the second hypothesis, whereby the change in reported sibship size with the respondent's age reflects past fertility patterns. Indeed, fertility increased in certain developing countries from the 1960s to the 1980s before starting to decline.

Preston's equivalence (1976) is useful for examining fertility trends in relation to reported sibship size. Preston showed that mean sibship size (denoted *F*, and including the reference individual) can be obtained without approximation as a function of the mean number of children ever born to the preceding generation (*G*) and the variance of this number (σ^2) such that

$$F = G + \frac{\sigma^2}{G}.$$

If all women had the same number of children, variance would be zero and sibship size would be equal to the number of children ever born to the previous generation. When variation is introduced into the parities, the mean sibship size is higher than the mean number of liveborn children. Indeed, larger sibships are mentioned more often, as they have a larger number of "representatives" in the population than smaller ones. Let us imagine a case where 100 women who have completed their reproductive life are distributed by parity as shown in Table 1. If mortality did not vary with family size (i.e. mother's completed parity), the probability of randomly drawing a mother who had 3 children would be 0.24, while that of randomly drawing a child who grew up in a family of 3 children would be 3 × 0.24. In this configuration, the mean number of children ever born would be 3.35, and the variance of this distribution 2.81, while the mean number of siblings would be 4.19, which corresponds to

$$3.35 + \frac{2.81}{3.35}.$$

By identifying the cohort of mothers who gave birth to the women interviewed in the DHS, it is thus possible to recalculate the "expected" sibship sizes of these women on the basis of their mothers' completed parity and its variance.

To identify the cohort of mothers, information is needed on the age difference between parents and children, i.e. the mean age at childbearing (MACB). This age varies little across countries or over time (between 25 and 30 years) and

Table 1. Calculating the mean number of siblings from the number of children ever born

	Number of children ever born									Total
	0	1	2	3	4	5	6	7	8	
Number of women	5	6	21	24	21	11	9	2	1	100

Number of children ever born (*G*)	$0.05 \times 0 + 0.06 \times 1 + 0.21 \times 2 + \ldots = 3.35$
Variance of number of children ever born (σ^2)	$(0.05 \times 0^2 + 0.06 \times 1^2 + 0.21 \times 2^2 + \ldots) - G^2 = 2.81$
Mean number of siblings (*F*)	$G + \frac{\sigma^2}{G} = 3.35 + \frac{2.81}{3.35} = 4.19$

tends to decrease during the fertility transition (Bongaarts, 1999). It is influenced primarily by two factors: postponement of births among the youngest mothers (following an increase in age at marriage, for example) and limiting of births among the oldest mothers. It can be obtained by weighting women's age by the number of births occurring at each age in the last 12 months.(1) For the countries analysed here, 530 estimates were gathered from DHS surveys and Multiple Indicator Cluster Surveys (MICS), from the United Nations Statistical Yearbook (United Nations, 1997; United Nations, 2009), from several other national demographic surveys and from census microdata in the IPUMS database.(2) These estimates are presented in the form of dots on Figure 2A. While there are numerous estimates for certain countries, such as Peru for example, others, such as Eritrea and Angola are much less well documented. To obtain a value for each country (j) and each year (t), a mixed linear model is used (Pinheiro et al., 2013). In this model (Equation 1), the constant and the rate of decrease of the mean age are allowed to vary by country:

$$\mathrm{MACB}_{ij} = \beta_0 + u_{0j} + \beta_1 t_{ij} + u_{1j} t_{ij} + e_{ij}$$

$$\text{with } u_{0j} \sim N(0, \sigma^2_{u0}),\ u_{1j} \sim N(0, \sigma^2_{u1}) \text{ and } e_{ij} \sim N(0, \sigma^2_e) \qquad (1)$$

The resulting predicted values will be close to the mean in countries for which estimates are scarce, or where they fluctuate substantially. They will "fit" better with the country estimates when such estimates are more numerous and more regular (Gelman and Hill, 2007).

The same model is used to reconstitute the trends in variance of completed parity. This variance is calculated on the basis of women aged 45-49, since this is the age group for which the largest amount of data is available (notably because the DHS surveys do not interview women aged above 50). We therefore

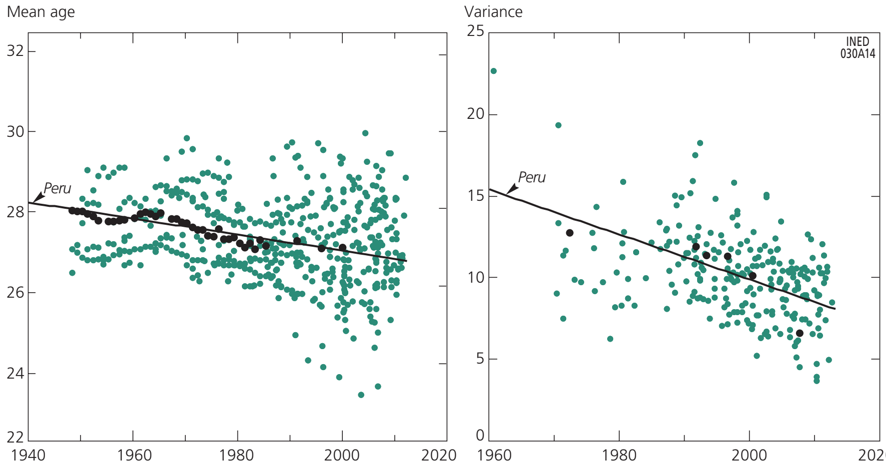
(1) It is often confused with "mean age of the fertility schedule" which is obtained by weighting mothers' ages by the fertility rates. The mean age at childbearing is influenced by the effect of mortality on the number of women. The difference between these two measures may be especially large in the case of a rapidly growing population.

(2) Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 6.1 [Machine-readable database]. Minneapolis, University of Minnesota, 2013.

assume that fertility above age 45 is marginal. Here again, the data are drawn from the DHS surveys, the IPUMS database and census reports or other national surveys. Like mean age at childbearing, variance in family size tends to decrease as fertility declines; this is visible on Figure 2B, in which Peru is again highlighted.

Figure 2. Trends in mean age at childbearing (A) and in variance in parity reached by women aged 45-49 (B) for 50 developing countries

A. Mean age at childbearing *B. Variance of parity at ages 45-49*



Sources: Demographic and Health Surveys; Multiple Indicator Cluster Surveys (http://www.childinfo.org/mics.html); United Nations Statistical Yearbooks (https://unstats.un.org/unsd/demographic/products/dyb/dyb2.htm); IPUMS database (https://international.ipums.org/international/).

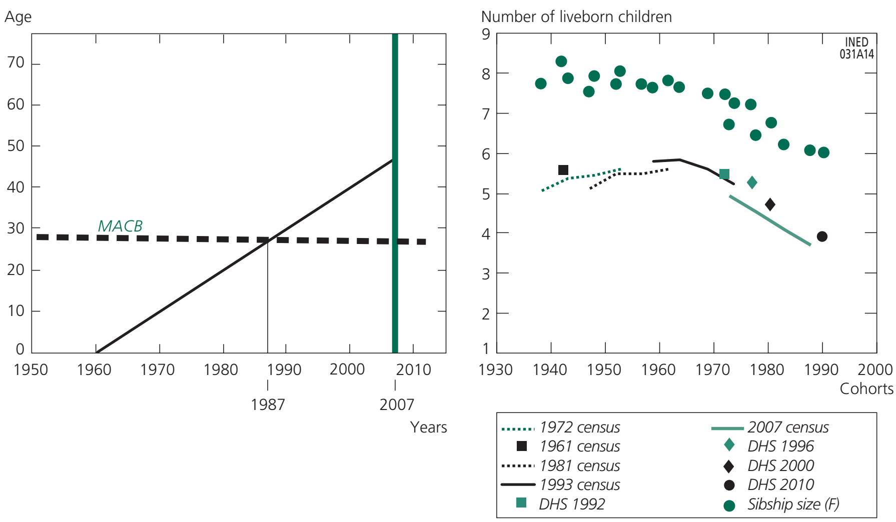
The data on the number of children ever born come from the same sources, but are more abundant, since certain survey or census reports publish mean parities without giving the full distribution. In all, 376 data collection operations are used: 175 DHS, 137 censuses and 64 other surveys (MICS, World Fertility Surveys, etc.). Only the completed parity of women in four ages groups were used (ages 45-49, 50-54, 55-59, and 60-64).

Why use completed parity as reference indicator when it may be of poor quality? It is generally acknowledged that older women tend to under-report their children (United Nations, 1983).(3) This idea emerged in the 1950s and 1960s, when it was found that completed parity increased more slowly with age than expected. It is nonetheless recognized today that temporary fertility

(3) For certain data collection operations, another problem arises when childless women are confused with women whose parity reached is unknown. This is the case, for example, if interviewers write a dash (-) in the corresponding questionnaire cell. El-Badry (1961) developed a method for correcting reported parities to take account of this problem. His method is applied here in cases where the distribution of women by age and parity was available (for example, with the IPUMS data), when the proportion of missing responses exceeded 2% and when the conditions necessary for this adjustment were met. The corrected values are also used to estimate the variance of the parities of women aged 45-49.

upturns occurred in several developing countries (Dyson and Murphy, 1985), which may partly explain why family size does not necessarily increase rapidly with women's age (Brass, 1996). Feeney (1991) has also shown that in several countries mean family size remains remarkably consistent from one collection operation to another. His method involves presenting completed parity on a graph, with the year in which the women reached mean age at childbearing plotted on the x-axis. Let us return to the example of Peru, where a census was held in 2007. As shown on Figure 3A, women aged 45-49 enumerated in this census were born, on average, in April 1960 (we assume that they are uniformly distributed across ages 45 to 50). In August 1987, they were aged 27.3 years, the mean age at childbearing in Peru estimated at that time. This is the date chosen as the reference period to present their parity.(4)

Figure 3. Example of reference period calculation (A), completed parities and expected sibship sizes in Peru (B)



Sources: Demographic and Health Surveys, United Nations Statistical Yearbooks, IPUMS database.

By proceeding in this way for the other three age groups up to 60-64, then for the other data collection operations in Peru, we obtain the pattern presented in Figure 3B. Note the high level of consistency of the parities reported over the period. Of course, it is possible that all parities are under-reported to a similar extent, but this overall consistency shows that omissions do not vary significantly with age. Combined with the variance of the number of children ever born presented above, these estimates can be used to calculate "expected"

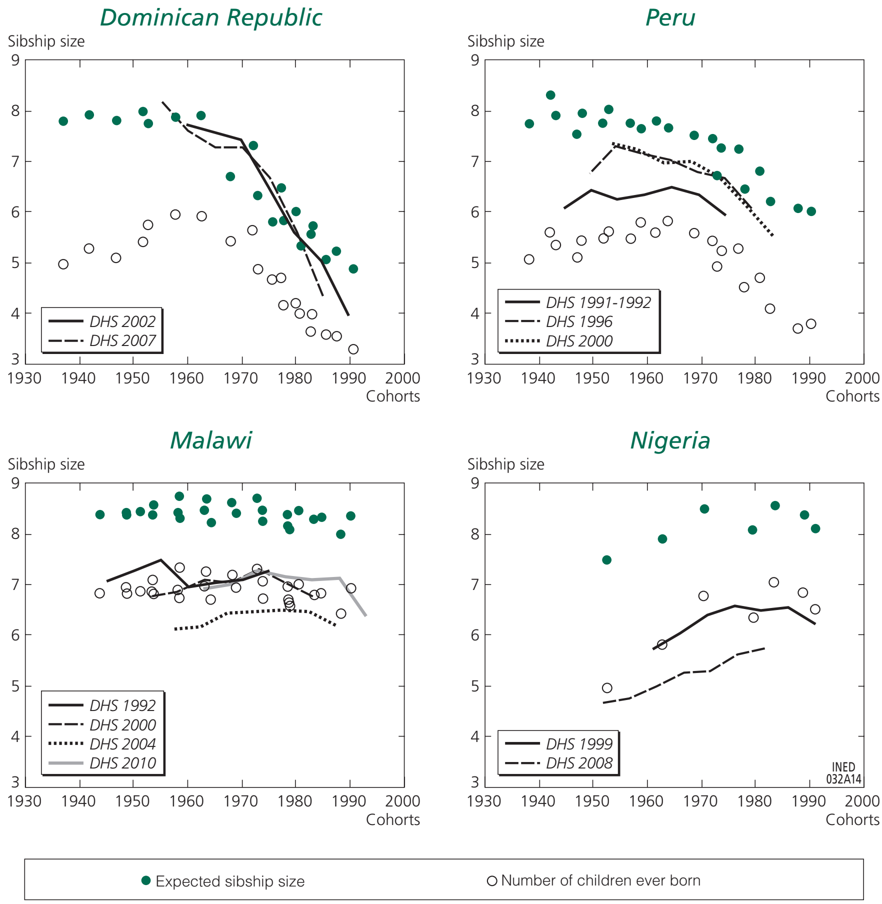
(4) Ideally, the mean age at childbearing should be calculated by cohort rather than by period, but the difference is marginal, given the low variability of this indicator.

family sizes (green dots). Expected size is around 8 persons on average for the cohorts born up to 1965 in Peru, then falls to 6 persons for the cohorts born around 1990.

II. Results

If correctly reported in the DHS, the sibships of adult women should be at least as large as those calculated from completed parity. However, Figure 4 shows that this is very rarely the case. It gives the completed parities (empty circles) and the expected sibship sizes (green dots) for four countries. The sibship sizes (including the respondent) reported in the DHS surveys are presented as continuous lines for various survey dates.

Figure 4. Numbers of children ever born, expected sibship sizes and sibship sizes reported in DHS surveys in Peru, Dominican Republic, Malawi and Nigeria



Sources: Demographic and Health Surveys, United Nations Statistical Yearbooks, IPUMS database.

Several configurations are observed. In the Dominican Republic, the reported and expected sibship sizes are very similar, and the two surveys are relatively concordant. This configuration, which reflects high-quality data, is observed in fewer than ten countries, most located in Asia or Latin America (they include Brazil, the Philippines, Nepal and Jordan). In Peru, the reported sibship sizes are lower than the expected sizes, but there are more reported siblings than children ever born. The mean difference is 1.25 siblings in the 1992 survey, versus less than 0.6 in the two following surveys. This situation is found in numerous countries, including Cambodia, Cameroon, Morocco and Uganda. The configuration observed in Malawi is the most frequent in sub-Saharan Africa: reported sibship sizes are below the expected sizes and closer to the number of children ever born. Last, in certain surveys, reported sibship sizes are much too small – smaller even than the number of children ever born, as is the case for surveys conducted in Nigeria in 1999 and Sierra Leone in 2008.

To quantify these differences, it is useful to interpolate between the various expected values. A cubic spline function is used here (Faraway, 2006). It thus becomes possible to calculate the relative difference (F1 – F2) / F1, where F1 is the expected sibship size and F2 is the size reported by adults in surveys. This difference can be interpreted as the proportion of siblings who were omitted, assuming that the numbers of children ever born are correctly reported. It comes to 16% on average across all surveys, and it varies by age: it is 14% among respondents aged 25-34, and reaches 20% among women aged 45-49. It is also slightly higher among the youngest respondents (17% at ages 15-19), probably because their sibship sizes are not yet final (even though their mothers, aged around 45, will have few additional children). Regional variations are more marked. For respondents aged 20 and above, the relative differences are 10% on average for surveys conducted in Latin America and the Caribbean versus 14% in southern Africa and 21% in West Africa. Table 2 lists the surveys by the scale of the relative difference between expected and reported sibship sizes. To measure whether these omissions disproportionately concern males or females, the relative differences between expected and reported sibship sizes are compared with the sex ratio at birth of brothers and sisters in Figure 5A. To obtain this sex ratio, the respondents were excluded to take account of the fact that only females were interviewed in the surveys used. If the majority of omissions concern sisters, the sex ratios should increase with the extent of the relative differences. This is doubtless what is observed in Afghanistan in 2010 and in Nigeria in 1999. But apart from these extremes, there is no clear association between sibling sex ratio and the extent of omissions. The correlation between these two indicators is low (<0.04) and non-significant. The mean sex ratio (1.04) is close to the theoretical value (1.05), although there is substantial variation across surveys, with half of surveys having a sex ratio below 1.01 or above 1.06. These variations probably reflect a genuine heterogeneity in sex ratios at birth, as demonstrated by Garenne (2002) for sub-Saharan Africa. The

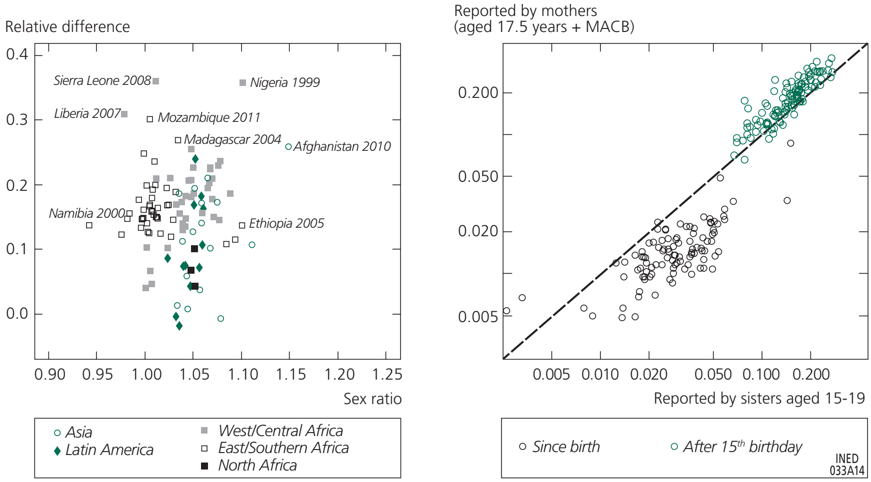
Table 2. List of DHS surveys including a module on siblings by extent of relative difference between expected and reported sibship sizes (respondents aged 20 and above)

	1990-1999	2000-2009	2010-2012
– 2% to + 4.9%	Brazil 1996 Jordan 1997 Nepal 1996 Philippines 1993, 1998	Congo 2005 Morocco 2003-2004 Dominican Republic 2002, 2007 São Tomé and Principe 2008-2009	
5% to 9.9%	Peru 1996 Sudan 1998-1999	Gabon 2000 Haiti 2000, 2005-2006 Nepal 2006 Peru 2000	Burundi 2010
10% to 14.9%	Morocco 1992 Namibia 1992 Togo 1998 Zambia 1996 Zimbabwe 1994, 1999	Bangladesh 2001 Cambodia 2005 Ethiopia 2000, 2005 Guatemala 1995 Indonesia 2007 Kenya 2003 Lesotho 2004 Namibia 2000, 2006-2007 Uganda 2000-2001, 2006 Rwanda 2005, 2010 Senegal 2005 Tanzania 2004-2005	Cambodia 2010 Cameroon 2011 Congo 2011-2012 Côte d'Ivoire 2011-2012 Ethiopia 2011 Gabon 2012 Indonesia 2012 Malawi 2010 Uganda 2011 Rwanda 2010 Senegal 2011
15% to 19.9%	South Africa 1998 Benin 1996 Indonesia 1994, 1997 Kenya 1998 Madagascar 1992, 1997 Malawi 1992 Uganda 1995 Peru 1991-1992 Senegal 1992-1993 Tanzania 1996	Benin 2006 Bolivia 2003, 2008 Cambodia 2000 Cameroon 2004 Congo (RDC) 2007 Côte d'Ivoire 2005 Ghana 2007 Guinea 2005 Indonesia 2002-2003 Kenya 2008-2009 Lesotho 2009 Madagascar 2008-2009 Malawi 2000 Mali 2006 Mozambique 2003 Rwanda 2000 Swaziland 2006-2007 Chad 2004 Zambia 2001-2002, 2007 Zimbabwe 2005-2006	Burkina Faso 2010 Tanzania 2010 Zimbabwe 2010-2011
20% to 24.9%	Bolivia 1994 Burkina Faso 1998-1999 Cameroon 1998 Central African Republic 1994-1995 Côte d'Ivoire 1994 Guinea 1999 Mali 1995-1996 Mozambique 1997 Chad 1996-1997	Burkina Faso 2003 Malawi 2004 Mali 2001 Niger 2006 Nigeria 2008	East Timor 2009-2010
25+%	Niger 1992 Nigeria 1999	Liberia 2007 Madagascar 2003-2004 Sierra Leone 2008	Afghanistan 2010 Mozambique 2011

**Figure 5. (A) Relationship between sibling sex ratios and relative difference between reported and expected sibship size;
(B) Proportions of deceased males reported by their sisters (aged 15-19) and by their mothers (aged 17.5 years + MACB) by age at death (since birth or after age 15)**

A. Relative difference in sibship size and sex ratio

B. Proportions of deceased males (log scale)



Source: Demographic and Health Surveys.

correlation between sex ratios at birth calculated from birth histories and from sibling data is relatively high (0.66).

Although the differences between expected and reported sibship sizes are large, it should be pointed out that sibling omissions do not necessarily result in underestimation of mortality, except if deceased siblings are omitted in a disproportionate manner. Here again, it is instructive to compare sisters' and mothers' reports. Within a given survey, respondents aged 15-19 provide information on cohorts of children born to mothers aged around 45, since mean age at childbearing is around 27 years. The proportion of deceased brothers can thus be compared, at aggregate level, with the proportion of deceased sons.[(5)] Only information on brothers and sons will be used here, since the respondents aged 15-19, all women, are also all survivors.

In Figure 5B, the proportion of deceased brothers as reported by their sisters aged 15-19 is plotted along the x-axis, and the proportion of sons as reported by mothers along the y-axis, distinguishing between deaths since birth or after the 15th birthday. The proportions of sons deceased are almost systematically higher than the proportions of deceased brothers. Yet these data come from the same surveys and refer to the same cohorts. In other words,

(5) It is necessary to interpolate between the proportions of deceased children of women aged 40-44 and of women aged 45-49 to obtain the corresponding value at 17.5 years plus the mean age at childbearing.

underestimation of mortality appears to be greater in the data on sibship size than in birth histories, probably because sisters tend to omit elder brothers who died before they were born. However, most of these deaths occurred in childhood, and when adult mortality is estimated biographically, these omissions do not bias the estimates. When only brothers who survived beyond their 15th birthday are used to calculate the proportion of deceased individuals, adult mortality appears to be reported more accurately by sisters than by mothers. These unexpected results call for further analysis. These differences may reflect variations in reporting accuracy, but also composition effects. Sibships which include a surviving sister aged 15-19 and at least one brother who survived beyond his 15th birthday may have faced risks of dying that differ from the overall experience of children born to mothers who are still alive (and aged around 45). Reporting errors could be identified by matching the reports of women aged 15-19 against their mother's birth history (although this is only possible when both mother and daughter were surveyed in the same household).

Conclusion

Comparison of sibship size and number of children ever born is based on several assumptions. In particular, we must assume that (1) the cohort of women identified from the mean age at childbearing is truly representative of the experience of all respondents' mothers; (2) there is no association between sibship size and mortality; and (3) women's completed parity is correctly reported. The first assumption could be refined using information on the respondents' birth order in order to better identify their mother's birth cohort, but this birth order is itself biased by sibling omissions, and we do not know whether these omissions concern the respondents' older or younger siblings. Moreover, published data on mean age at childbearing are rarely distributed by parity. The other two assumptions are more problematic, but they tend to cancel each other out. First, although there is little association between adult survival and number of adult siblings, child mortality, on the other hand, generally increases with the mother's parity because of competition between the children for resources and parental attention, inadequate living space, physical exhaustion of the mother and transmission of infections (Zaba and David, 1996). As a consequence, there are fewer adult survivors in large sibships, leading to a larger difference between expected and reported sibship sizes. Conversely, the under-reporting of children ever born (United Nations, 1983) tends to narrow the gap between expected and reported sibship sizes, suggesting that omissions may be more frequent than indicated here. It is difficult at this stage to assess how the violation of these two assumptions affects the comparisons of sibship sizes and numbers of children ever born. Simulations will be run to take the analysis further.

Despite these limits, comparing sibship data with mean parities is a more reliable approach than simply examining the number of siblings in isolation. In the DHS survey reports, sibship size is regularly used as an indicator of data quality. If sibship size does not increase monotonously with the respondents' age, then omissions are suspected. The analysis presented here shows that this indicator is too crude. Given that the sibships were created several decades before the survey, their size should not necessarily increase with age. Not only must surveys be compared against each other, but additional data on mothers' parity can also be used to reveal the substantial under-reporting of siblings (15% on average). Stanton et al. (2000) had assumed that the older female respondents omitted more siblings, and this is confirmed, although the differences are small. Our analysis shows rather that omissions concern all age groups. These omissions are more frequent in sub-Saharan Africa than in other developing regions, notably in West Africa, perhaps because families are both larger and more complex in that region due to high fertility and widespread polygamy. Polygamy is generally associated with a high frequency of union dissolution. Age differences between spouses are also large. This factor, combined with high male mortality, results in shorter union durations and frequent remarriage. Consequently, the majority of women marry more than once, making it more likely for respondents to omit siblings born to the same mother but to different fathers, even though the DHS surveys ask respondents to report all siblings born to the same biological mother, including those with a different father. In their validation study in southern Senegal, Helleringer et al. (2014) indeed found that the data on sisters born to different fathers were of lesser quality, with the omission of 18% of half-sisters in a survey similar to DHS, versus just 7% of sisters with the same father and mother. The fact that in patrilinear societies children tend to stay with their father or his family after parental separation may partly account for these differences. But it should be noted that in polygamous societies, there are far fewer half-siblings with the same mother than with the same father. For example, in his study of the Peul Bandé in western Senegal, Pison (1986) estimated that in the early 1980s, the mean number of siblings with the same mother and father was 6.5, with a further 4.9 half-siblings having the same father but a different mother, and 0.9 half-siblings having the same mother but a different father. The inclusion by respondents of siblings with the same father but a different mother may also bias mortality estimates if the deaths of these half-siblings are less accurately reported. However, no analyses have been performed to date to determine whether reporting quality varies according to whether or not the siblings have the same mother.

Child circulation is also frequent in sub-Saharan Africa, and children are sometimes "fostered" out to others. This may also lead to reporting errors. For example, according to the latest DHS surveys in sub-Saharan Africa, more than one in four young people aged 10-14 whose mother is still alive do not live with her (www.statcompiler.com). This proportion ranges from an average of

21% in East Africa to 35% in southern Africa. By comparison, it is 16% in the surveys conducted in Latin America and the Caribbean, and around 5% in Asia. One can assume that individuals who did not live with their mother throughout their childhood will provide less reliable information on their siblings and will themselves be omitted by their brothers and sisters. Questionnaires could be modified to avoid such errors, notably by adding certain control questions, for example: "Are there any other brothers or sisters who were born a long time before you or after you and who you did not live with for very long?"

What are the implications of these findings for estimating adult mortality? There are currently three approaches for estimating mortality from sibling data. Hill and Trussell (1977) first developed an indirect method, comparable to Brass's method for child survival (Brass, 1996). It involves converting proportions of deceased siblings into probabilities of dying since birth. As the method covers the entire sibship, it is more sensitive to the omissions highlighted here and should therefore be used solely in cases where no data on siblings' ages have been recorded. It is preferable to use the indirect method developed by Timaeus et al. (2001) which applies to surveys or censuses in which a few additional questions are asked about the number of siblings who survived up to their 15th birthday. As omissions appear to mainly concern siblings who died in childhood, this method will produce less biased mortality levels. It does mix recent and older deaths, however, making it sensitive to other errors, such as under-reporting of adult deaths in the distant past, which appears to be a particularly worrisome problem (Masquelier et al., 2014). Ideally, a third approach should be preferred, involving direct analysis of sibling data (Rutenberg and Sullivan, 1991) so that analysis can focus on the most recent periods even if, here again, omissions inevitably still occur. However, in order to apply this biographical method, questions on the siblings' current age or their age at death must be asked. These questions exist in the DHS but not in the MICS surveys. In regions where coverage of vital registration is inadequate and only a small proportion of deaths are registered, this form of data collection should be used in all national surveys on health and mortality. Indeed, despite manifest errors and omissions, sibling survival data contribute greatly to our understanding of adult mortality levels and trends in developing countries.

Acknowledgements: I would like to thank Patrick Gerland who provided me with estimates on parities of women aged over 45 in several countries of sub-Saharan Africa, as well as John B. Casterline, Gilles Pison and three anonymous reviewers for their constructive comments. This research was performed under a research grant from the Fonds national de la recherche scientifique belge (FNRS).

References

Bendavid Eran, Holmes Charles, Bhattacharya Jay, Miller Grant, 2012, "HIV development assistance and adult mortality in Africa", JAMA, 307(19), pp. 2060-2067.

Bongaarts John, 1999, "The fertility impact of changes in the timing of childbearing in the developing world", *Population Studies*, 53(3), pp. 277-289.

Brass William, 1996, "Demographic data analysis in less developed countries: 1946-1996", *Population Studies*, 50(3), pp. 451-467.

Dyson Tim, Murphy Mike, 1985, "The onset of fertility transition", *Population and Development Review*, 11(3), pp. 399-440.

El-Badry Mohamed A., 1961, "Failure of enumerators to make entries of zero: errors in recording childless cases in population censuses", *Journal of the American Statistical Association*, 56(296), pp. 909-924.

Faraway Julian J., 2006, *Extending the Linear Model with R*, Taylor & Francis, 301 p.

Feeney Griffith, 1991, "Child survivorship estimation: Methods and data analysis", *Asian and Pacific Population Forum*, 5(2-5), pp. 51-55, pp. 76-87.

Gakidou Emmanuela, King Gary, 2006, "Death by survey: estimating adult mortality without selection bias from sibling survival data", *Demography*, 43(3), pp. 569-585.

Gakidou Emmanuela, Hogan Margaret, Lopez Alan D., 2004, "Adult mortality: Time for a reappraisal", *International Journal of Epidemiology*, 33(4), pp. 710-717.

Garenne Michel, 2002, "Sex ratios at birth in African populations: a review of survey data", *Human Biology*, 74(6), pp. 889-900.

Gelman Andrew, Hill Jennifer, 2007, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, 625 p.

Hagopian Amy, Flaxman Abraham D., Takaro Tim K., Esa Al Shatari Sahar, Rajaratnam Julie et al., 2013, "Mortality in Iraq associated with the 2003–2011 war and occupation: Findings from a national cluster sample survey by the University Collaborative Iraq Mortality Study", *PLoS Medicine*, 10(10), e1001533.

Helleringer Stéphane, Pison Gilles, Kanté Almamy M., Duthé Géraldine, Andro Armelle, 2014, "Reporting errors in siblings' survival histories and their impact on adult mortality estimates: results from a record linkage study in Senegal", *Demography*, 51(2), pp. 387-411.

Helleringer Stéphane, Duthé Géraldine, Kanté Almamy M., Andro Armelle, Sokhna Cheikh et al., 2013, "Misclassification of pregnancy-related deaths in adult mortality surveys: case study in Senegal", *Tropical Medicine & International Health*, 18(1), pp. 27-34.

Hill Kenneth, Trussell James, 1977, "Further developments in indirect mortality estimation", *Population Studies*, 31(2), pp. 313-334.

Masquelier Bruno, 2013, "Adult mortality from sibling survival data: a reappraisal of selection biases", *Demography*, 50(1), pp. 207-228.

Masquelier Bruno, Reniers Georges, Pison Gilles, 2014, “Divergences in trends in child and adult mortality in sub-Saharan Africa: survey evidence on the survival of children and siblings”, *Population Studies*, 68(2), pp. 161-177.

Merdad Leena, Hill Kenneth, Graham Wendy, 2013, “Improving the measurement of maternal mortality: the sisterhood method revisited”, *PLoS ONE*, 8, e59834.

Obermeyer Ziad, Rajaratnam Julie K., Park Chang H., Gakidou Emmanuela, Hogan Margaret C. et al., 2010, “Measuring adult mortality using sibling survival: A new analytical method and new results for 44 countries, 1974-2006”, *PLoS Medicine*, 7(4), e1000260.

Pinheiro José, Bates Douglas, DebRoy Saikat, Sarkar Deepayan, R Core Team, 2013, NLME: *Linear and Nonlinear Mixed Effects Models*, R package version 3.1-109.

Pison Gilles, 1986, “La démographie de la polygamie”, *Population*, 41(1), pp. 93-122.

Preston Samuel H., 1976, “Family sizes of children and family sizes of women”, *Demography*, 13(1), pp. 105-114.

Reniers Georges, Masquelier Bruno, Gerland Patrick, 2011, “Adult Mortality in Africa”, in Rogers Richard G., Crimmins Eileen M. (eds.), *International Handbook of Adult Mortality*, Springer, pp. 151-170.

Rutenberg Naomi, Sullivan Jeremiah M., 1991, “Direct and indirect estimates of maternal mortality from the sisterhood method”, *Demographic and Health Surveys World Conference Proceedings*, August 5-7, Washington, DC. Volume III. Calverton, Maryland USA, IRD/Macro International Inc. pp. 1669–1696.

Stanton Cynthia, Noureddine Abderrahim, Hill Kenneth, 2000, “An assessment of DHS maternal mortality indicators”, *Studies in Family Planning*, 31(2), pp. 111-123.

Timaeus Ian M., Jasseh Momodou, 2004, “Adult mortality in sub-Saharan Africa: evidence from demographic and health survey”, *Demography*, 41(4), pp. 757-772.

Timaeus Ian M., Zaba Basia, Ali Mohamed, 2001, “Estimation of adult mortality from data on adult siblings”, in Zaba Basia, Blacker John (eds.), *Brass Tacks: Essays in Medical Demography*, Athlone Press, London, pp. 43-66.

United Nations, 1983, *Manual X: Indirect Techniques for Demographic Estimation*, Population Division, Department of International Economic and Social Affairs, United Nations, New York.

United Nations, 1997, *Demographic Yearbook 1997 - Historical supplement*, Department of International Economic and Social Affairs, United Nations, New York.

United Nations, 2009, *Demographic Yearbook 2007*, Department of International Economic and Social Affairs, United Nations, New York.

Wang Haidong, Dwyer-Lindgren Laura, Lofgren Katherine T., Rajaratnam Julie K., Marcus Jacob et al., 2012, “Age-specific and sex-specific mortality in 187 countries, 1970-2010: a systematic analysis for the Global Burden of Disease Study 2010”, *Lancet*, 380(9859), pp. 2071-2094.

Wilmoth John R., Mizoguchi Nobuko, Oestergaard Mikkel, Say Lale, Mathers Colin et al., 2012, “A new method for deriving global estimates of maternal mortality”, *Statistics, Politics and Policy*, 3(2), pp. 1-38.

Zaba Basia, David Patricia H., 1996, “Fertility and the distribution of child mortality risk among women: an illustrative analysis”, *Population Studies*, 50(2), pp. 263-278.

Bruno MASQUELIER • SIBSHIP SIZES AND FAMILY SIZES IN SURVEY DATA USED TO ESTIMATE MORTALITY

Survey data on sibling survival provide a crucial source of information for estimating adult mortality in countries where vital records are incomplete. This article assesses the quality of these data by comparing sibship sizes reported in Demographic and Health Surveys with women's mean number of children ever born in the previous generation. This comparison, conducted at aggregate level, suggests that a high proportion of siblings are omitted, since the sibship sizes are 15% lower, on average, than would be expected on the basis of number of children ever born. Such omissions are more frequent in sub-Saharan Africa than in other developing regions, and their extent increases slightly with the respondents' age. Adult mortality deduced from these data is not necessarily underestimated, however, since omissions appear to mainly concern siblings who died in childhood.

Bruno MASQUELIER • TAILLE DES FRATRIES ET TAILLE DES FAMILLES DANS LES DONNÉES D'ENQUÊTES UTILISÉES POUR ESTIMER LA MORTALITÉ

Les données d'enquêtes recueillies sur la survie des frères et sœurs constituent une source incontournable pour estimer la mortalité des adultes dans les pays où l'état civil reste incomplet. Cet article évalue la qualité de ces données en comparant la taille des fratries déclarées dans les enquêtes démographiques et de santé avec le nombre moyen d'enfants nés vivants des femmes de la génération précédente. Cette comparaison, menée au niveau agrégé, suggère qu'une proportion élevée de frères et sœurs sont omis; les tailles de fratries sont inférieures de 15 % environ aux tailles attendues sur la base des enfants nés vivants. Ces omissions sont plus fréquentes en Afrique subsaharienne que dans les autres régions en développement et leur ampleur augmente légèrement avec l'âge des enquêtées. La mortalité aux âges adultes déduite de ces données n'est pas pour autant sous-estimée, car les omissions semblent surtout concerner des frères et sœurs décédés dans l'enfance.

Bruno Masquelier • TAMAÑO DE LAS FRATRÍAS Y TAMAÑO DE LAS FAMILIAS EN LOS DATOS DE ENCUESTA UTILIZADOS PARA ESTIMAR LA MORTALIDAD

Los datos de encuesta recogidos sobre la supervivencia de los hermanos y hermanas constituyen una fuente necesaria para estimar la mortalidad de los adultos en los países con un estado civil incompleto. Este artículo evalúa la calidad de esos datos comparando el tamaño de las fratrías declaradas en las encuestas demográficas y de salud con el número medio de hijos nacidos vivos de mujeres de la generación precedente. Esta comparación, hecha a nivel agregado, sugiere que una proporción importante de hermanos y hermanas es omitida; los tamaños de las fratrías son un 15% inferiores à los que cabría esperar sobre la base de los niños nacidos vivos. Estas omisiones son más frecuentes en África subsahariana que en las otras regiones en desarrollo y su amplitud aumenta ligeramente con la edad de las mujeres encuestadas. Sin embargo, la mortalidad a la edad adulta deducida de estos datos no está subestimada pues las omisiones parecen afectar sobre todo a los hermanos y hermanas muertos durante la infancia.

Keywords: Adult mortality, data quality, children ever born, mortality estimation, omissions, Demographic and Health Surveys

Translated by Catriona Dutreuilh.