Improved analysis of sibling survival data taking into account survivor bias, zero-surviving reporters and recall bias

Alison Levin-Rector, Julie Knoll Rajaratnam, Haidong Wang, Austin Schumacher, Carly E. Levitz, Christopher J.L. Murray

Abstract

Background:

In the absence of a well-functioning vital registration system to track mortality in a population, health planners often rely on routine health surveys to provide this most basic health information. Sibling survival histories, where a survey respondent is asked about each of his or her siblings' births and, if applicable, deaths, provide a direct way to estimate adult mortality by survey. The purpose of this paper is to refine the methods which account for the selection bias, zero-surviving reporters and recall bias inherent in these surveys to generate plausible estimates of adult mortality risks even in the presence of a relationship between family size and adult mortality.

Methods:

We have implemented changes to the previous method, referred to as the Corrected Sibling Survival (CSS) method, such that it (1) uses appropriate survival weights that account for the study design, and (2) recovers the mortality experience of the families that are not represented because none of the siblings is alive and eligible to respond to the survey. We validate these methodological developments in a range of simulation environments. We also present new ways of adjusting for recall bias and handling sparse data in survey designs where the age range of the respondents is narrower than the age range desired for estimation. We apply these methods to sibling history data from a number of sources and compare the agreement between the estimates from sibling histories and other sources of adult mortality data where available.

Results:

Simulation results demonstrate that the method generates unbiased estimates of adult mortality when there is zero association between family size and mortality; the estimates are too high in the presence of positive association, and too low in the

presence of negative association, although the bias is only statistically significant at the highest positive levels of association observed in actual survey data. When compared to estimates of adult mortality from independent sources of mortality data, estimates from sibling history surveys are comparable.

Conclusions:

Surveys with sibling histories can be used to generate plausible estimates of summary indicators of adult mortality and provide valuable information on the levels and trends of adult mortality in settings where vital registration systems do not exist.

Background

In the absence of a well-functioning vital registration system to track mortality in a population, health planners often rely on routine health surveys to provide this most basic health information. As death is increasingly postponed to adult and older ages, it is becoming even more important to be able to reliably, and cost-effectively, measure levels and trends in adult mortality. Yet for the majority of low and middle income countries, vital registration systems that register births and deaths in populations are still very incomplete, and it may well take decades before they can be used for monitoring mortality. Sibling survival histories, where a survey respondent is asked about each of his or her siblings' births and, if applicable, deaths, provide a direct and more immediate way to estimate adult mortality from surveys, pending improvements to vital registration systems. A cross-sectional survey not only provides current estimates of population-level mortality risks; it can also provide information about mortality trends 10-15 years ago.

Despite the critical importance of reliable information on adult mortality, enthusiasm for using sibling histories to estimate levels and trends of adult mortality has been tempered by concerns of selection and recall bias, as well as general data quality concerns.[1-5] For more than 20 years, researchers have attempted to address some of the shortcomings of sibling history data and to extract more accurate estimates of adult mortality.[2-4,6,7] This paper builds upon previous methodological research to address key biases in sibling history data and presents an improved method which is less reliant upon assumptions than methods commonly used in current practice.

A major source of bias in sibling history data is recall, specifically omission, bias. Comparison of results across multiple surveys from the same country suggests that respondents tend to under-report deaths of siblings which have occurred in the distant past.[2,3,5] Statistical solutions to adjust for recall bias are thus essential if sibling history based data are to be useful for guiding policy and programs to reduce premature adult mortality. Recent research has demonstrated that such solutions are possible.[3]

Another primary source of bias is selection bias. If mortality rates are computed directly from the data (referred to as the 'naïve' estimator by Gakidou and King)[7], the fact that the respondent (usually a woman) is alive at the time of the survey will bias the estimates downwards. One way of addressing this bias is to exclude the respondent from the denominator in the calculation of the rates. Trussell and Rodriguez show that under conditions where mortality does not vary with the size of the family of siblings, known as sibship size, this method generates estimates of mortality that are unbiased.[6] However, given a strong likelihood that, at least in some cases, mortality is in fact related to sibship size,[4,7] there is a need for a method that does not depend on whether or not this assumption holds true.

A further desirable feature for any method to analyse sibling history data is that it will be able to generate estimates of 45q15, the probability that, under current mortality conditions, a 15 year old will die before reaching his or her 60th birthday. 45q15 is a standard indicator of premature adult mortality and widely used as a summary indicator of population health.[1,2,8-12] Because data from the DHS are collected only from respondents within the age range of 15-49, computing an estimate of 45q15 presents a challenge, because the number of sibling records above age 50 is fewer, and the data become even more sparse for periods of time more distant from the date of the survey. A

variety of solutions have been proposed to address this limitation in the DHS. Timaeus & Jasseh imposed an age pattern on the mortality data based on models developed from historical mortality patterns along with a model to adjust these age patterns when necessary due to the HIV/AIDS epidemic.[2] Others have employed variations of this approach.[4,9] Although the HIV/AIDS component of their model allows deviation from historical patterns, questions remain about the extent to which these historical patterns represent contemporary experience in developing country populations.[13] In a different approach, Obermeyer and colleagues estimated four age patterns of mortality by pooling together the DHS sibling history data.[3] The benefit of this approach is that the age patterns are defined by the data themselves; however, it is not clear how generalizable the mortality patterns over time and across all 49 countries which have conducted a DHS sibling history survey will be. This is a subject of further research.

The purpose of this paper is to refine existing methods that account for selection bias to generate plausible estimates of adult mortality rates even in the presence of a relationship between adult sibship size and mortality. We have implemented changes to the method developed by Obermeyer and colleagues[3], referred to as the CSS method, such that it (1) uses appropriate survival weights that account for the study design (as described by Masquelier)[4], and (2) recovers the mortality experience of the sibships that are not represented in the sampling frame either because all siblings have died or because there are no siblings alive who are eligible to participate in the survey (e.g. no women between ages 15-49 in the case of DHS). We also present new ways of adjusting for recall bias and handling sparse data in survey designs where the age range of the respondents is narrower than the age range desired for estimation. We apply these methods to sibling history data from a number of sources and compare the findings to estimates derived from other sources of adult mortality data where available.

Methods

Data:

In this paper, we have used simulated data, data from the DHS, the Centers for Disease Control Reproductive Health Surveys (CDC-RHS), and data from surveys conducted in four field sites (Bohol, Philippines; West Godavari, AP, India; Shivgarh, UP, India; and Pemba, Tanzania) by the Population Health Metrics Research Consortium (PHMRC).[14-16] In all of these surveys, a sibling history (also known as a maternal mortality module) collects information from the respondent about each sibling born to the same biological mother, including their sex, age, whether alive, and if dead, the year of death. Annex Table 1 describes each of these survey families and summarizes the characteristics of the sibling history modules from the surveys within them.

Analytic methods:

Gakidou-King weights to adjust for selection bias

Selection bias refers to the underrepresentation of high mortality sibships in the sample population—sibships with higher rates of mortality are less likely to be represented in the survey because fewer of them are likely to have survived to be selected into the sample. A method to correct for this underrepresentation, proposed by Gakidou and King,[7] incorporates a sibship-level weight, $W_j = B_j/S_j$, where B_j is the original sibship size and S_j is the number of siblings in sibship j who survive to the time of the survey. When each observation in the dataset being analysed is at the sibship level, this Gakidou-King (GK) weight can be used to compute a weighted average of the

proportions of siblings deceased as reported by each respondent. In the absence of any sibships where all siblings have died, this correction algebraically corrects for the underrepresentation of high-mortality sibships in the survey sample.

When the dataset is expanded to the sibling level (i.e. one observation for each sibling as opposed to sibship), the number of observations listed in the dataset for each sibship corresponds to the original sibship size, B_j, and so the numerator of W_j is already accounted for. The resulting sibling-level weight is therefore, $W_i = 1/S_j$ for sibling i in sibship j.[4,9] Since the analysis reported here is carried out at the sibling level, we use W_i rather than W_j. This improves on previous applications of the method where the sibship-level weight was inappropriately applied to data that had been expanded to the sibling level.[3]

Further, the number of surviving siblings in the family must also be tailored to the eligibility criteria for respondents of the given survey.[4] In applying Gakidou and King's elucidation of the survivorship correction, S_j/B_j represents the probability that a sibling in sibship j survives and is eligible to be selected in the survey. For DHS surveys, respondents must be women between the ages of 15 and 49 and so the S_j in this case would be the number of surviving women in sibship j who are between the ages of 15 and 49 at the time of the survey. In this analysis, the value of S_j has been chosen to be consistent with the eligibility criteria of each survey.

Zero-reporter correction

Another important source of selection bias is zero-reporter bias. This is the bias that arises because the sampled portion of the population, by definition, excludes sibships in which there are no siblings that are eligible to respond to the survey and therefore report on the mortality experience of their siblings. This bias will be larger in populations with higher levels of mortality. In the case of the DHS, the population not represented in the sample includes sibships that contain only men, or some combination of men, women outside of the 15 to 49 age range, and/or women who would be within this age range but have died. For the cases of male mortality and for female mortality outside the 15 to 49 age range, we assume that the mortality rates in the omitted population are the same as those of the siblings represented by respondents in the sample, and so are unbiased. Mortality rates within the age range of eligible reporters, unless they account for siblings of women who would have been eligible to report had they not died before the time of the survey, will be biased downward. Thus, we need to further adjust the mortality rates of siblings that are within the age and sex eligibility criteria to account for this bias.

An updated correction for this bias from previous methods[3] directly estimates the number of sibling deaths that are missing from the sample by age and sibship size and subsequently adds these siblings to the observed sample before calculating age-specific mortality risks. Since smaller sibships are more likely to be zero-reporter sibships, we begin by applying this method to sibships with one or two females (for surveys that interview both men and women, the method is applied to sibships with one or two total siblings.) A detailed explanation of this correction is provided in the Appendix.

Compared to the number of missing women estimated for single-sister sibships, we observed between 3% and 14% more missing women after applying the correction to two-sister sibships among all surveys. Thus, the relative contribution that would be gained by expanding to the three-sister case is likely very small and would not contribute appreciably to the estimated mortality rates. Given the required

computational efforts to expand to the three-sister case, we did not go beyond the two-sister (or two-sibling) case.

Validation of Methods to Adjust for Survivor Bias
In order to validate and test the performance of the updated methods to correct for selection bias as described above, we used a micro-simulation approach to create an environment where true mortality rates are known. We generated a dataset with individuals who then are exposed over time to age-specific risks of mortality and fertility. We assume no migration in this case.

We modelled populations using data from five countries in order to simulate heterogeneous mortality and fertility schedules. We imposed the age-distribution of the population of Sweden in 1751 on the initial population and applied constant single year age- and sex-specific probabilities of death and single year age-specific fertility rates for 96 years to achieve a stable population. We then applied 61 years of observed and variable age- and sex- specific fertility rates extracted from the Human Fertility Database (HFD), as well as age and sex- specific mortality rates from model life tables developed in the Global Burden of Disease and Risk Factors 2010 study, for the years 1946 through 2006 from Canada, Finland, Sweden, Switzerland, and the USA.[17,18]

We further expanded the test environments to include scenarios where we imposed an association between sibship size and mortality. We accomplished this by multiplying single year probabilities of death by a scalar that depended on the sibship size of each sibling. Four additional scenarios were created: strong positive association, strong negative association, and the maximum level of positive and negative association observed in the DHS. Strong positive association was imposed by multiplying the age- and sex-specific probabilities of death by half of the sibship size. Strong negative association was imposed by multiplying the age- and sex-specific probabilities of death by the sibship size of the index sibling subtracted from the maximum sibship size in the population all multiplied by ½. For example, if the maximum sibship size in the population was 16, the probability of death for a sibling in a sibship of size 10 would be multiplied by a factor of (16 – 10) * ½, or 3. The probability of death for a sibling in a sibship of size 2 would be multiplied by a factor of (16 – 2) * ½, or 7. In this manner, probabilities of death for siblings in smaller sibships were inflated more than probabilities of death for siblings in larger sibships. We set the multiplier for the maximum sibship size to be ½.

In order to determine the maximum level of association that existed between sibship size and mortality in the DHS, we first calculated 35q15 for males by sibship size within each survey with only female respondents. After dropping outlying sibship sizes on a per-survey basis, we ran a regression that relates the log difference between sibship-size-specific 35q15 and overall 35q15 to sibship size as a categorical variable. We then exponentiated the regression coefficients and smoothed them by regressing on sibship size as a continuous variable. The resulting regression equation gives us the linear formula to find the sibship-specific scalars that we multiply to the overall age- and sex-specific probability of death to get the sibship-size-specific probability of death. Slopes of the scalars by sibsize are -0.084 and 0.25 for maximum negative and positive associations found in DHS respectively. The multipliers can be found in Table 3. We chose 35q15 as our summary measure because it is less influenced by smaller numbers of respondents at the older ages than is 45q15, and we chose to look at males because they are not affected by zero-reporter bias.

In all four scenarios, after multiplying the age- and sex-specific probability of death by the sibship-size-specific scalar, we multiplied these values by the ratio of the age- and

sex-specific mean of the original probability of death to the age- and sex-specific mean of the association-adjusted probability of death. This serves to recapture the overall levels of input mortality to ensure we are not inflating/deflating mortality rates.

In total, we generated 30 unique populations for each of the five scenarios depicting levels of association between sibship size and mortality risk. Within each of these 150 modelled populations, we simulated 100 surveys, sampling approximately 10,000 respondents in the sampling scheme of the DHS, where women between the ages of 15 and 49 are eligible to report, as well as in the sampling scheme of the PHMRC, where both men and women aged 15 or older are eligible to respond. Assuming no misreporting, we gathered data from respondents, replicating observations in cases where more than one sibling from a sibship was selected as a respondent, and dropping sibships in which no one was selected. We then estimated age-specific mortality rates by applying our methods to the survey data to assess how the results compared to the true mortality rates calculated from the entire simulated population. Uncertainty intervals were generated for each scenario by pooling estimates from the 100 surveys of each of the 30 populations for each scenario and calculating the 2.5th and 97.5th percentiles from these 3,000 estimates.

Addressing the issue of sparse data to generate estimates of 45q15
In DHS-type surveys, which limit the ages of respondents to between 15 and 49, the estimates of mortality in the age range from 50-59 are derived solely from the experience of older siblings of respondents. Therefore, the numbers of siblings and sibling deaths in these age groups is less than in age groups that include respondents. The data become even more sparse further back in time, since the cohorts of respondents and their siblings cover an even younger age range in the past. This can generate small numbers problems if the data are too sparse. To address this issue, we used the following logistic regression model:

$$logit(Y_{at} = 1) = \beta_0 + \beta I_a + \beta I_t + \varepsilon$$

where Y_{at} is an indicator of survival or death in age group a at time t, I_a is a set of indicators for each five year age group, 15-19, 20-24, … , 55-59, and I_t is a set of indicators for each five year period prior to the year of the survey. In using this model we are assuming a constant age pattern over the fifteen years prior to the survey, effectively borrowing strength across all periods. In addition, we analyse each survey separately, which means that the level and patterns observed in the underlying data are preserved to a greater extent than would be the case if age patterns were imposed or data pooled across surveys and countries.

The resulting age patterns may still be "noisy", especially in surveys with a smaller sample size and/or lower mortality. We therefore conducted a sensitivity analysis to assess how the quality of the age pattern affects the level of 45q15. To do this, we smoothed the age patterns obtained from the logistic regression models using kernel-weighted local polynomial smoothing and compared the resulting estimates of 45q15 to the 45q15s generated from the age patterns as directly observed from the regression results. In addition, we recognize that the assumption of a constant age pattern over a fifteen year period is tenuous in the context of major epidemiologic change such as that brought about by HIV. To determine how assuming a constant age pattern affects estimates of 45q15, we performed an additional sensitivity test by selecting all surveys from Sub-Saharan Africa that showed a monotonic increase in adult mortality consistent with what we would expect in a setting with a high prevalence of HIV. We compared estimates of 45q15 obtained by separately analysing data from the five year period

immediately prior to the survey (the period where we would expect a clear HIV-related age pattern) to estimates of 45q15 obtained by pooling data from this period with the two periods 5-9 and 10-14 years prior to the survey (when HIV prevalence was much lower) to see how much the assumption of a constant age pattern over time affects the level of the resulting summary-level indicator of interest, namely 45q15.

Recall bias adjustment
To quantify and adjust for recall bias in the sibling history data, we combined the estimates of 45q15 from all women's surveys, and aligned the estimates for all periods where they overlapped. Since we estimated mortality levels for three five-year periods prior to the survey year, estimates overlapped when at least two surveys were carried out in the same country less than 15 years apart. For example, analysis of the surveys in Burkina Faso yields two estimates for the five-year period 1993-1997; one estimate is generated from the 2003 survey and the other estimate is from the 1998 survey. There were 298 unique pairs of overlapping estimates; 152 for females and 146 for males. The discrepancy by sex is due to the fact that the CDC-RHS surveys only collected information about sisters of respondents, so male mortality could not be ascertained from this survey family.

The estimates of mortality for each five-year period were assigned to the midpoint of the period, and the period of recall for each estimate was calculated by subtracting from the year that the survey was conducted. In the Burkina Faso example, the average period of recall for the estimate from the 2003 survey would be 7.5 years while the average period of recall for the estimate from the 1998 survey would be just 2.5 years. Thus the difference in the period of recall between the estimates from these two surveys was 5 years. We calculated this difference in years of recall between each pair of points as well as the magnitude of the difference between the two estimates of 45q15 and implemented a linear regression model separately by sex of sibling to quantify the relationship between years of recall and the level of mortality. The coefficient on recall period represents the effect of recall bias and was then used to adjust the estimates accordingly.

Comparison to other sources of mortality
We compared the estimates of 45q15 generated from applying the methods outlined above to all surveys in countries or settings for which alternate, independent sources of adult mortality data were available. These sources included vital registration data (evaluated and adjusted for incompleteness if necessary),[21] and in the case of the data from the PHMRC sites, the sources included estimates from a dual-census comparison and identification of deaths through tracing (Pemba, Shivgarh and West Godavari) and a capture-recapture analysis of three systems for recording deaths (Bohol).[22]

Results

Validation of survival bias correction in simulation environments
The population distribution by age and sex for the simulated populations is shown in Figure 1.

Figure 2 shows the application of the components of the survival bias correction to a sibling history survey in each association population. When both the GK weights and the zero-reporter corrections are applied, the method for adjusting for survival bias recovers the underlying true rates better than applying each component separately.

Table 1 presents the estimates of relative error of 45q15 with associated uncertainty intervals for each of the scenarios of association between sibship size and mortality for the DHS- and PHMRC-type sampling schemes. Table 2 presents mean residuals with standard deviations in estimated 45q15 among each simulation scenario and the differences between the Corrected Sibling Survival method and the Trussell-Rodriguez method.

In the DHS-type sampling scheme, with no association between sibship size and mortality, estimates are unbiased, with the relative error of estimated 45q15 differing by -0.3% (95%UI: -9.7%, 8.8%) for females and -1.4% (-7.7%, 5.0%) for males. Under strong negative association, the estimates are biased downwards by -9.1% (-17.4%, -0.7%) for females and -8.2% (-14.1%, -2.1%) for males. Under strong positive association, the estimates are biased upwards, by 11.5% (2.7%, 22.0%) for females and 14.7% (8.1%, 21.7%) for males. The bias in scenarios modelled according to the level of association observed in the DHS is markedly less; the estimates are biased downward for the maximum negative association by -3.4% (-12.1%, 5.3%) for females and -4.4% (-10.3%, 1.8%) for males, while the estimates for the maximum positive association observed in the DHS is biased upward by 7.2% (-1.5%, 16.9%) for females and 8.4% (2.1%, 14.5%) for males. The PHMRC-type sampling scheme seems to display downward biases over each of the association scenarios, but the corresponding uncertainty intervals are much wider than those in the DHS sampling scheme.

In comparing our method to the Trussell-Rodriguez method, mean residuals of the 45q15 estimate are comparable for no association between sibship size and mortality (female: 0.0002 vs. -0.0008; male: 0.0015 vs. 0.0015), maximum DHS negative association (female: 0.002 vs. 0.002; male: 0.005 vs. 0.006), and strong negative association (female: 0.005 vs. 0.009; male: 0.007 vs. 0.012). The method proposed in this article performs better for the maximum DHS positive association (female: -0.004 vs. -0.008; male: -0.009 vs. -0.013) and strong positive association (female: -0.007 vs. -0.012; male: -0.016 vs. -0.022). For the PHMRC sampling scheme, results are more sporadic and overall comparable.

Application of the methods to sibling history surveys and comparison to other sources of mortality data

For males, the all-country coefficient on recall was 0.00429, meaning that the level of 45q15 is expected to decrease by this amount for each additional year prior to the survey. For females, the annual decrease was lower at 0.00440. We used these coefficients to adjust estimates to account for recall bias in addition to selection bias and zero-reporter bias. The impact of these three components of our method is illustrated for six surveys in Figure 3. The GK weights raise the estimated 45q15 by 5.9%; the addition of the zero-reporter siblings raise the estimated 45q15 by 9.5%; and the addition of the recall bias correction further raise the estimates by 16.7%, on average.

Results of the application of our methods to sibling history surveys at the country-level are shown in Appendix Figure 1, along with the estimates of 45q15 from other available sources for comparison. While the estimates from the sibling histories can be fairly different compared to those derived from other data sources, an overall comparison (Figure 4) suggests an average consistency (in other words, there is definitely "noise" in the relationship but no apparent bias).

At a more detailed level, we observe as expected that analysing individual surveys can result in particularly erratic estimates of mortality by age. However, after smoothing the

age patterns for all survey-periods, the mean absolute change in 45q15 was 1.6% (range: 0.0%, 12%), indicating a robust estimation of 45q15, despite "noisy" age patterns. Similarly, the effect of assuming a constant age pattern over time among countries with steep increases in HIV prevalence versus using the pattern observed in the raw data also yields robust estimates of 45q15 (a comparison of 45q15 using both approaches shows a mean absolute change of 2.4% with a range from 0.5% to 6.3%).

Discussion

We have presented a method which can be applied to sibling history data from retrospective surveys to yield estimates of adult mortality that are broadly consistent with other, generally more reliable sources. In particular, our method appears to adequately address the problem of selection bias as demonstrated in a variety of micro-simulation environments. In scenarios of extreme association between sibship size and mortality, the method produces point estimates that are biased, but under more plausible levels of association, the accompanying uncertainty intervals include the comparator value (i.e. the true level of adult mortality).

The biases which are evident when strong associations between sibship size and mortality risk are imposed are likely due to the fact that the probability of being a zero-reporter sibship depends in part on the size of the sibship. Regardless of mortality, smaller sibships are more likely to be completely missed in our sample; therefore their mortality experience will be underrepresented. In the case of positive association between sibship size and mortality risk, this results in an overestimation of mortality because we are differentially missing smaller sibships which, by definition, experience a lower risk of mortality. In the case of negative association, the resulting estimate is conversely downwardly biased. While our method does not account for this aspect of zero-reporter bias, our exploration of the resulting bias when we imposed the maximum levels of positive and negative association observed in the DHS showed that this association only leads to significant bias in our estimates of 45q15 for the maximum positive association when using the DHS-type sampling scheme. The maximum positive association is very extreme as compared to the rest of the DHS surveys analysed, and using associations more in the normal range of the surveys does not lead to significant bias. With the PHMRC-type sampling scheme, we get uncertainty intervals that are very wide, making any inference difficult. To a certain degree, this is due to the much wider range of ages sampled—any individual over age 15—which ends up providing less information on the 15-60 age group than the DHS-type sampling scheme, thus we have more uncertainty for 45q15. This is especially true for populations with older age distributions such as those in our micro-simulations.
The comparison of our estimates with estimates derived from independent sources of mortality data suggests an unclear relationship between different sources, but no obvious indication that the sibling estimates are biased. Given the amount of uncertainty in both sources (the uncertainty intervals in the sibling history estimates are generally quite wide as shown in the graphs in Appendix Figure 1, and adjusting vital registration data for incompleteness can add substantial uncertainty to the comparison estimates),[21] this "noisiness" is not surprising.

One major limitation of this analysis is the estimation of the average recall bias across all surveys and the use of this average effect in calculating levels and trends in 45q15. Because of the scarcity of countries in which multiple surveys have been conducted within 15 years of each other, using the average recall bias estimate is necessary. As more countries conduct more surveys that include sibling survival modules, it will

become possible to apply this method on a regional and eventually perhaps even country-by-country basis. More specific recall bias parameter values can then be used to generate estimated levels of adult mortality.

An additional limitation of this method is the inability to rely on the estimated age patterns of mortality for further inference; however as we have shown, the implausibility of the age patterns does not appreciably affect the summary indicator of 45q15. The assumption of a constant age pattern over time also is not likely; however, this assumption has only a small effect on the summary indicator of interest, namely 45q15.

These limitations need to be understood, but they are, in our view, by no means the reason for not utilizing sibling history technique. Rather, the methods presented here now permit improved estimation of adult mortality in a number of countries where information gaps exist because of the lack of adequate systems to register deaths. As more countries collect sibling survival data, it will be possible to explore the contextual, linguistic and other cultural factors that might account for variability in recall bias. This knowledge could help to guide further improvements in survey instruments for recalling sibling deaths.

Measuring adult mortality more reliably in poor populations is becoming an increasingly urgent public health priority as global health attention shifts to include an emphasis on prevention and control of non-communicable diseases. This ought to be accompanied by growing recognition of the potential utility of sibling history data for public health monitoring.

References

[1] Gakidou E, Hogan M, Lopez AD. Adult mortality: time for a reappraisal. *International Journal of Epidemiology*. 2004;33(4):710 -717.

[2] Timaeus I, Jasseh M. Adult Mortality in Sub-Saharan Africa: Evidence From Demographic and Health Surveys. *Demography*. 2004;41(4):757-772.

[3] Obermeyer Z, Rajaratnam JK, Park CH, et al. Measuring Adult Mortality Using Sibling Survival: A New Analytical Method and New Results for 44 Countries, 1974–2006. *PLoS Med*. 2010;7(4):e1000260.

[4] Masquelier B. Adult mortality from sibling survival data: a reappraisal of selection biases. *forthcoming. Demography*.

[5] Stanton C, Abderrahim N, Hill K. An Assessment of DHS Maternal Mortality Indicators. *Studies in Family Planning*. 2000;31(2):111-123.

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

[7] Gakidou EE, King G. Death by Survey: Estimating Adult Mortality Without Selection Bias From Sibling Survival Data. *Demography*. 2006;43(3):569-585.

[8] Rajaratnam JK, Marcus JR, Levin-Rector A, et al. Worldwide mortality in men and women aged 15-59 years from 1970 to 2010: a systematic analysis. *The Lancet*. 2010;375(9727):1704-1720.

[9] Reniers G, Masquelier B, Gerland P. Adult Mortality in Africa. In: Rogers RG, Crimmins EM, eds. *International Handbook of Adult Mortality*.Vol 2. Dordrecht: Springer Netherlands; 2011:151-170. Available at: http://www.springerlink.com/content/g82222m300072147/. Accessed July 26, 2011.

[10] World Bank. *Disease and mortality in Sub-Saharan Africa*. World Bank Publications; 2006.

[11] Blacker J. The impact of AIDS on adult mortality: evidence from national and regional statistics. *AIDS*. 2004;18 Suppl 2:S19-26.

[12] Jahn A, Floyd S, Crampin AC, et al. Population-level effect of HIV on adult mortality and early evidence of reversal after introduction of antiretroviral therapy in Malawi. *The Lancet*. 2008;371(9624):1603-1611.

[13] INDEPTH Network. *INDEPTH model life tables for Sub-Saharan Africa*. Ashgate Publishing, Ltd.; 2004.

[14] ICF Macro. Demographic and Health Surveys: MEASURE DHS Home. Available at: http://measuredhs.com/. Accessed July 26, 2011.

[15] Centers for Disease Control and Prevention. Surveys: Global Reproductive Health | CDC Reproductive Health. Available at: http://www.cdc.gov/reproductivehealth/Surveys/index.htm. Accessed July 26, 2011.

[16] Population Health Metrics Research Consortium. Population Health Metrics Research Consortium | Institute for Health Metrics and Evaluation. Available at: http://www.healthmetricsandevaluation.org/research/project/population-health-metrics-research-consortium. Accessed July 26, 2011.

[17] Max Planck Institute for Demographic Research, Vienna Institute of Demography. Human Fertility Database. Available at: http://www.humanfertility.org/cgi-bin/main.php. Accessed July 26, 2011.

[18] Wang H, Dwyer-Lindgren L, Lofgren KT, et al. Age-specific and sex-specific mortality in 187 countries, 1970—2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; **380**: 2071-2094.

[19] Wang H, Lopez AD, Murray CJL. Estimating age-specific mortality: a new model life table system with flexible standard mortality schedule. *IHME Working Paper*. 2011.

[20] US Census Bureau DIS. International Programs, Database. Available at: http://www.census.gov/population/international/data/idb/informationGateway.php. Accessed July 27, 2011.

[21] Murray CJL, Rajaratnam JK, Marcus J, Laakso T, Lopez AD. What Can We Conclude from Death Registration? Improved Methods for Evaluating Completeness. *PLoS Med*. 2010;7(4):e1000262.

[22] Carter KL, Williams G, Tallo V, et al. Capture-recapture analysis of all-cause mortality data in Bohol, Philippines. *Popul Health Metr*. 9:9-9.

Tables

Table 1: Relative error in estimated ${}_{45}q_{15}$ among various simulation scenarios

Sampling Scheme	Level of Association Between Sibship Size and ${}_{45}q_{15}$	Mean Relative error in estimated ${}_{45}q_{15}$ with 95% UI (%)	
		Female	Male
DHS	No association	-0.3 [-9.7,8.8]	-1.4 [-7.7,5.0]
DHS	Max DHS Negative association	-3.4 [-12.1,5.3]	-4.4 [-10.3,1.8]
DHS	Max DHS Positive association	7.2 [-1.5,16.9]	8.4 [2.1,14.5]
DHS	Strong negative association	-9.1 [-17.4,-0.7]	-8.2 [-14.1,-2.1]
DHS	Strong positive association	11.5 [2.7,22.0]	14.7 [8.1,21.7]
PHMRC	No association	-2.2 [-37.3,41.8]	-8.1 [-30.1,14.7]
PHMRC	Max DHS Negative association	-0.9 [-33.5,40.8]	-5.0 [-25.1,17.3]
PHMRC	Max DHS Positive association	-3.2 [-37.6,41.3]	-9.5 [-31.5,14.4]
PHMRC	Strong negative association	-0.3 [-29.6,38.5]	-3.9 [-24.6,18.0]
PHMRC	Strong positive association	-0.3 [-9.7,8.8]	-1.4 [-7.7,5.0]

Table 2: Mean (standard deviation) of residuals in estimated ${}_{45}q_{15}$ among various simulation scenarios and the differences between the Corrected Sibling Survival method and the Trussell-Rodriguez method

		Mean residual in estimated ${}_{45}q_{15}$ with standard deviation			
		CSS		Trussell-Rodriguez	
Sampling Scheme	Level of Association Between Sibship Size and ${}_{45}q_{15}$	Female	Male	Female	Male
DHS	No association	0.0002 (0.003)	0.0015 (0.003)	-0.0008 (0.003)	0.0015 (0.004)
DHS	Max DHS Negative association	0.002 (0.003)	0.005 (0.003)	0.002 (.003)	0.006 (0.004)
DHS	Max DHS Positive association	-0.004 (0.003)	-0.009 (0.003)	-0.008 (.003)	-0.013 (0.004)
DHS	Strong negative association	0.005 (0.003)	0.009 (0.003)	0.007 (0.003)	0.012 (0.003)
DHS	Strong positive association	-0.007 (0.003)	-0.016 (0.004)	-0.012 (0.004)	-0.022 (0.004)
PHMRC	No association	0.0013 (0.011)	0.0081 (0.013)	-0.0003 (0.008)	-0.0001 (0.011)
PHMRC	Max DHS Negative association	0.001 (0.012)	0.009 (0.012)	0.003 (0.008)	0.005 (0.010)
PHMRC	Max DHS Positive association	0.001 (0.011)	0.005 (0.012)	-0.008 (0.008)	-0.014 (0.011)
PHMRC	Strong negative association	0.002 (0.012)	0.010 (0.013)	0.007 (0.008)	0.011 (0.010)
PHMRC	Strong positive association	0.0002 (0.010)	0.004 (0.012)	-0.012 (0.009)	-0.022 (0.011)

Table 3: Multipliers for the DHS maximum positive and maximum negative association between sibship size and mortality. In the mortality and sibship size association equation, x = sibsize.

Sibship Size	Association multiplier	
	Maximum Negative *1.54 – 0.084x*	Maximum Positive *0.499 + 0.25x*
1	1.456	0.749
2	1.372	0.999
3	1.288	1.249
4	1.204	1.499
5	1.120	1.749
6	1.036	1.999
7	0.952	2.249
8	0.868	2.499
9	0.784	2.749
10	0.700	2.999
11	0.616	3.249
12	0.532	3.499

Figures

Figure 1: Population distributions by age and sex for the simulation environments with (a) zero association between mortality and sibship size, (b) the maximum DHS negative association, (c) the maximum DHS positive association, (d) strong negative association, and (e) strong positive association.

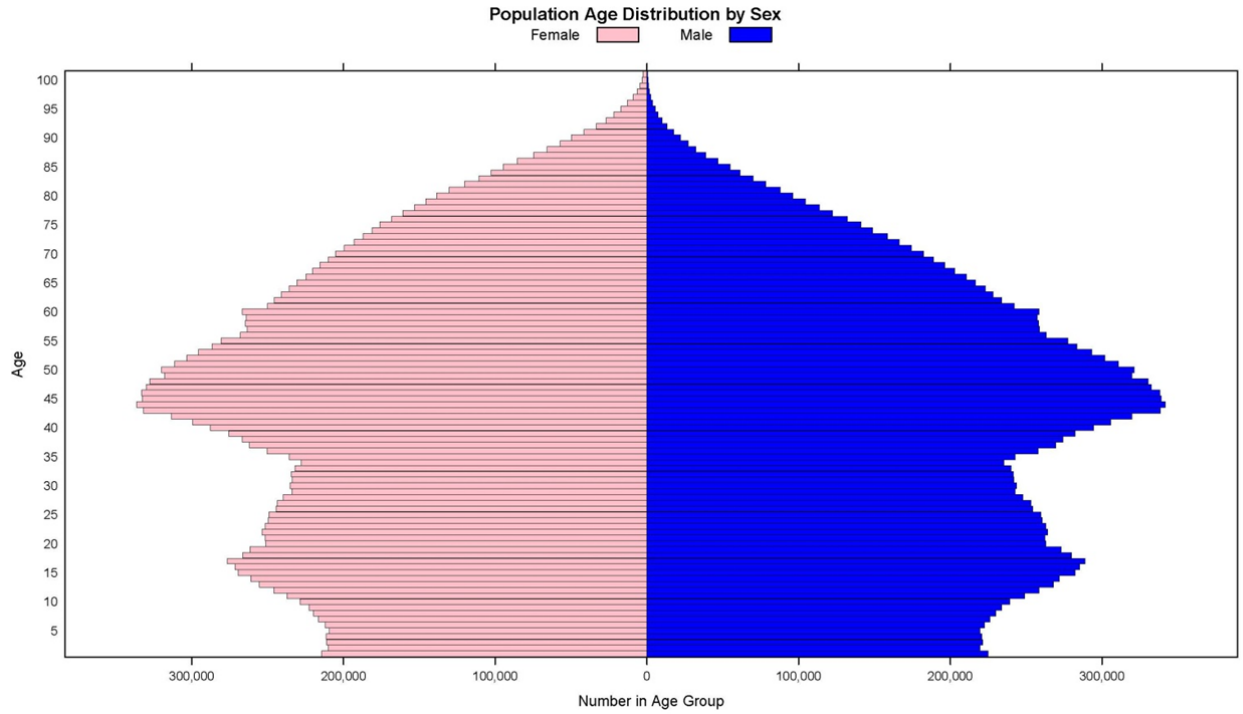
Figure 2: Age-specific mortality rates for the five years prior to the survey. Step-by-step results when applying the DHS sampling strategy to low mortality simulated population, with (a) zero association between sibship size and mortality, (b) the maximum DHS negative association, (c) the maximum DHS positive association, (d) strong negative association, and (e) strong positive association.

Figure 3: A step-by-step look at each of the adjustments in the Sibling Survival Method for a selection of DHS: Cambodia 2000, Ethiopia 2005, Haiti 2005, Malawi 2000, Niger 1992, and Philippines 1998.

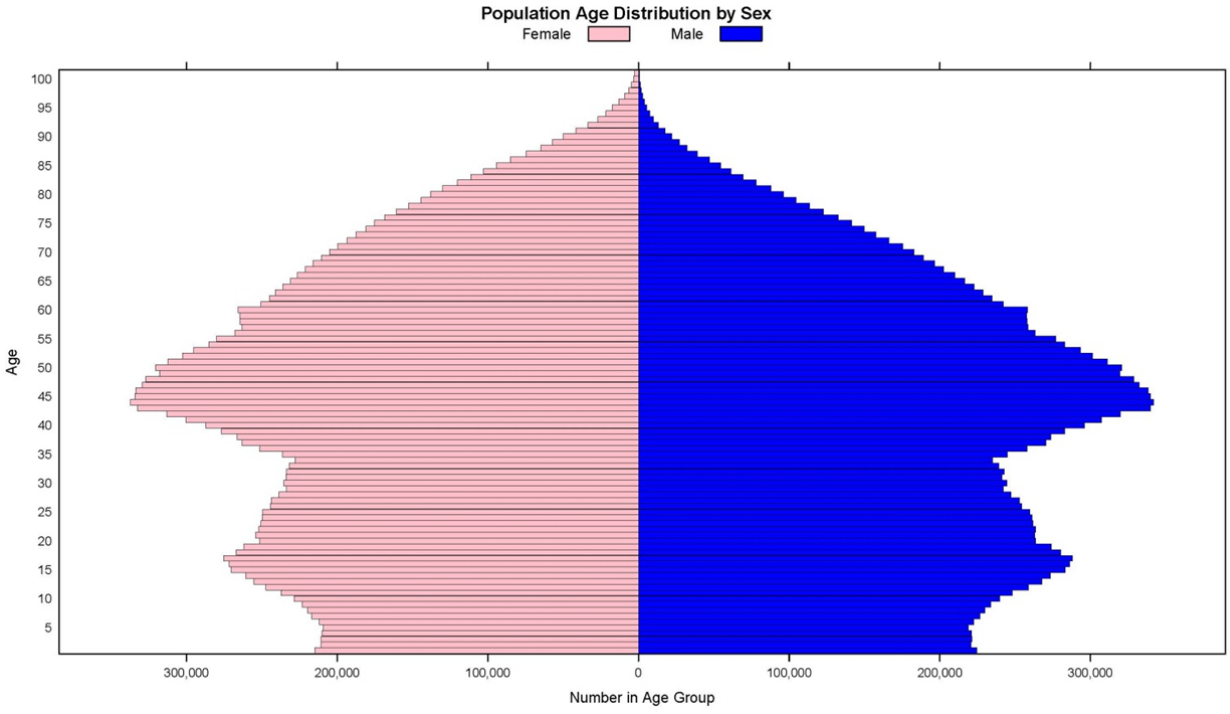
Figure 4: Comparison of estimates of 45q15 generated from sibling history surveys to estimates of 45q15 from other independent sources of adult mortality data.

Figure 1

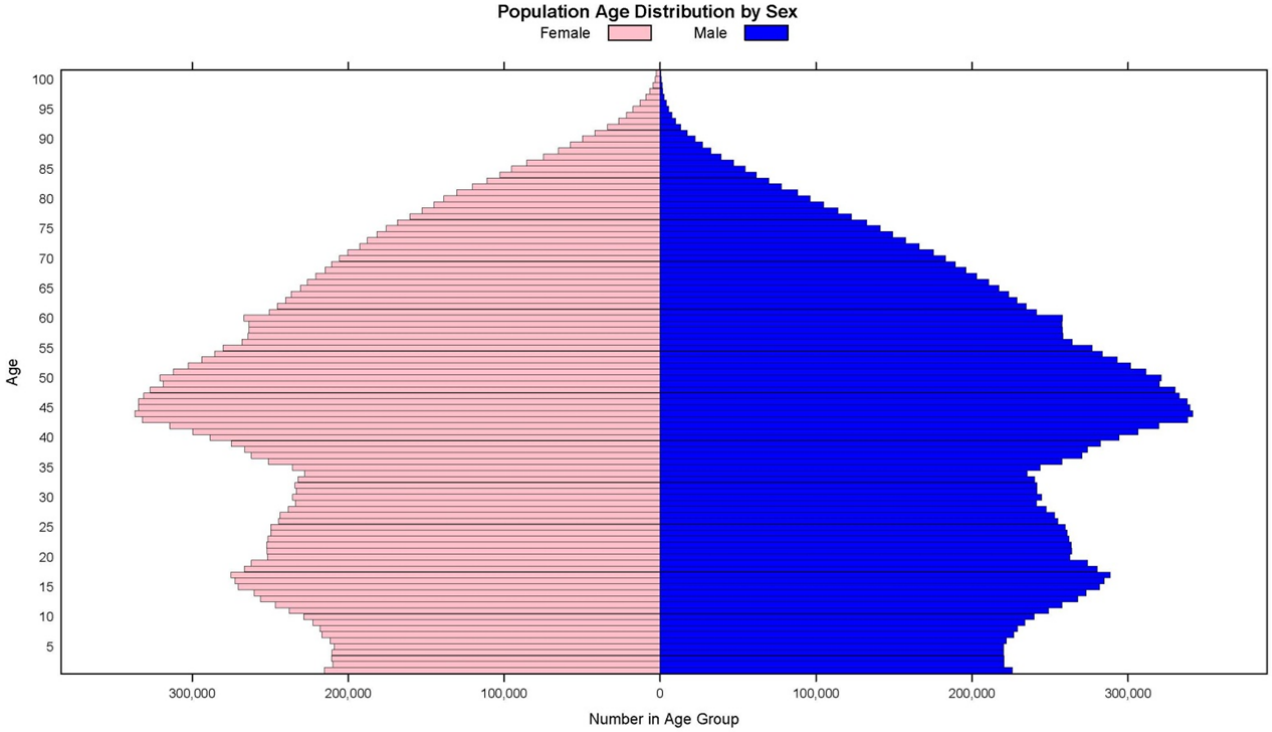
(a)


Population Age Distribution by Sex
Female
Male
Age
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
300,000
200,000
100,000
0
100,000
200,000
300,000
Number in Age Group

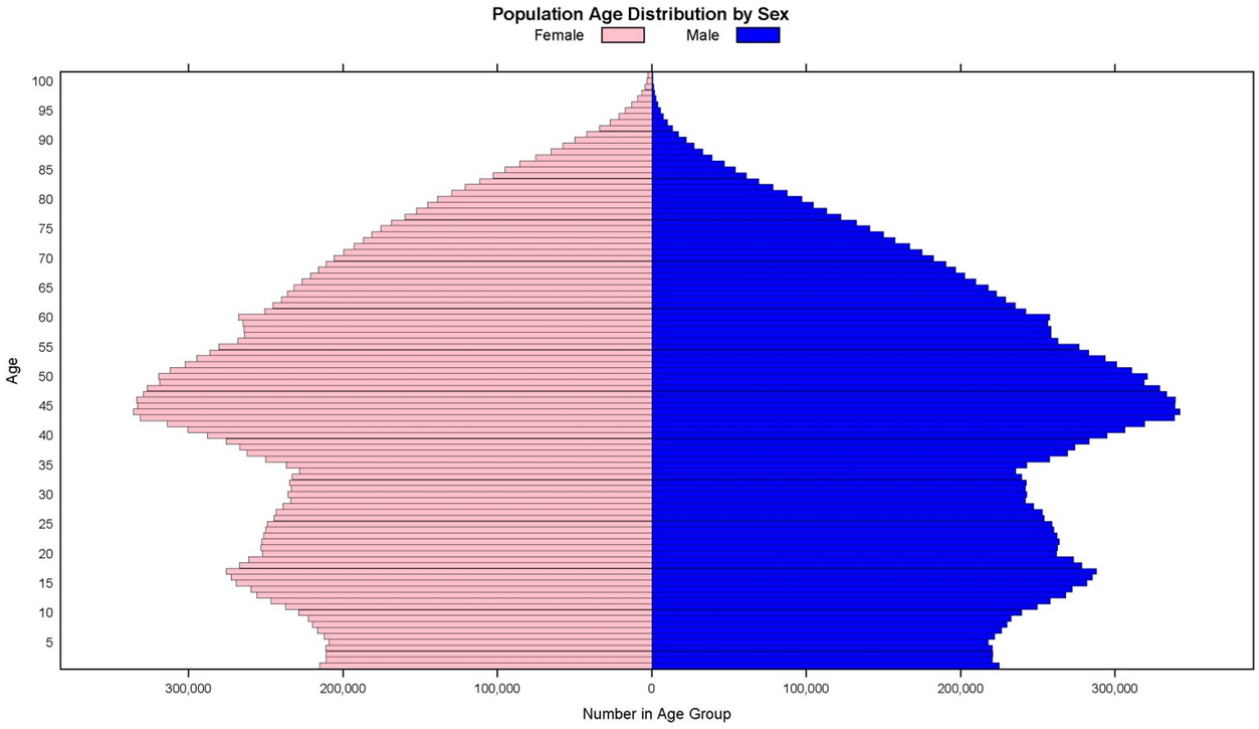
(b)


Population Age Distribution by Sex
Female
Male
Age
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
300,000
200,000
100,000
0
100,000
200,000
300,000
Number in Age Group

(c)


Population Age Distribution by Sex
Female
Male
Age
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
300,000
200,000
100,000
0
100,000
200,000
300,000
Number in Age Group

(d)


Population Age Distribution by Sex
Female
Male
Age
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
300,000
200,000
100,000
0
100,000
200,000
300,000
Number in Age Group

(e)

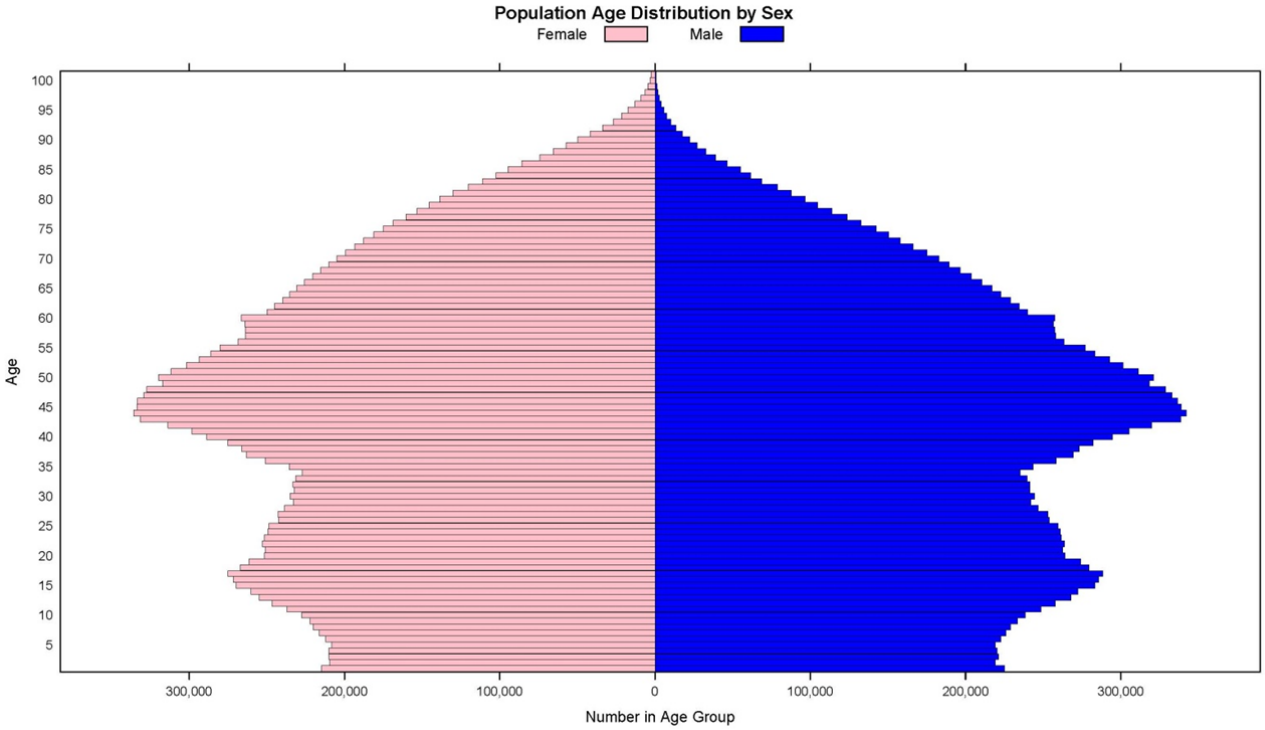
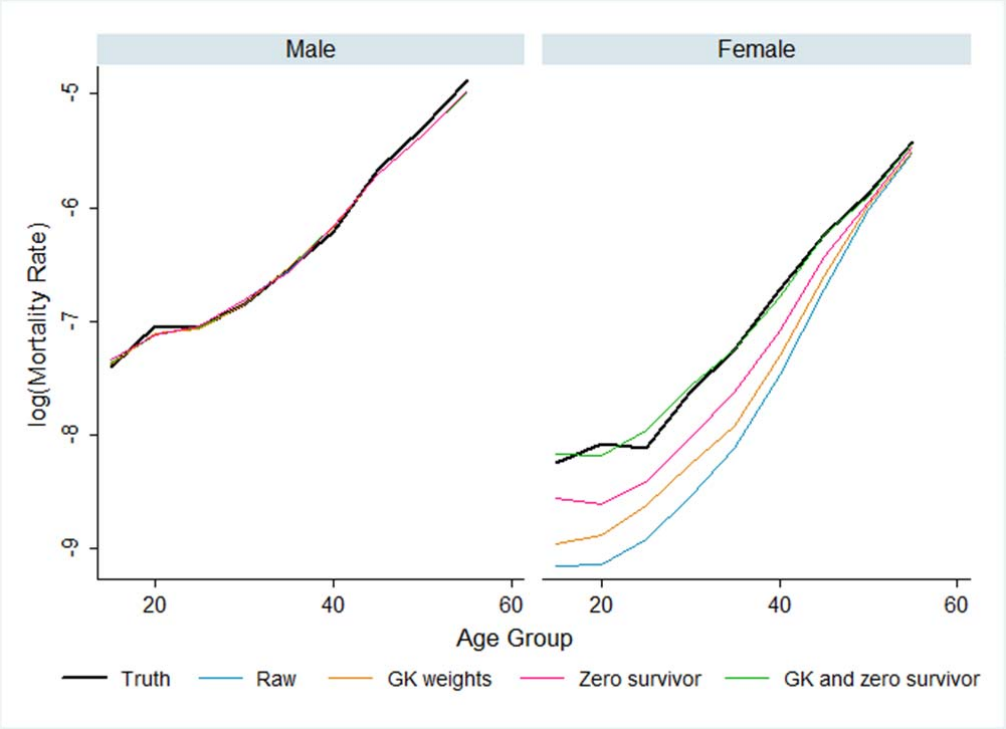
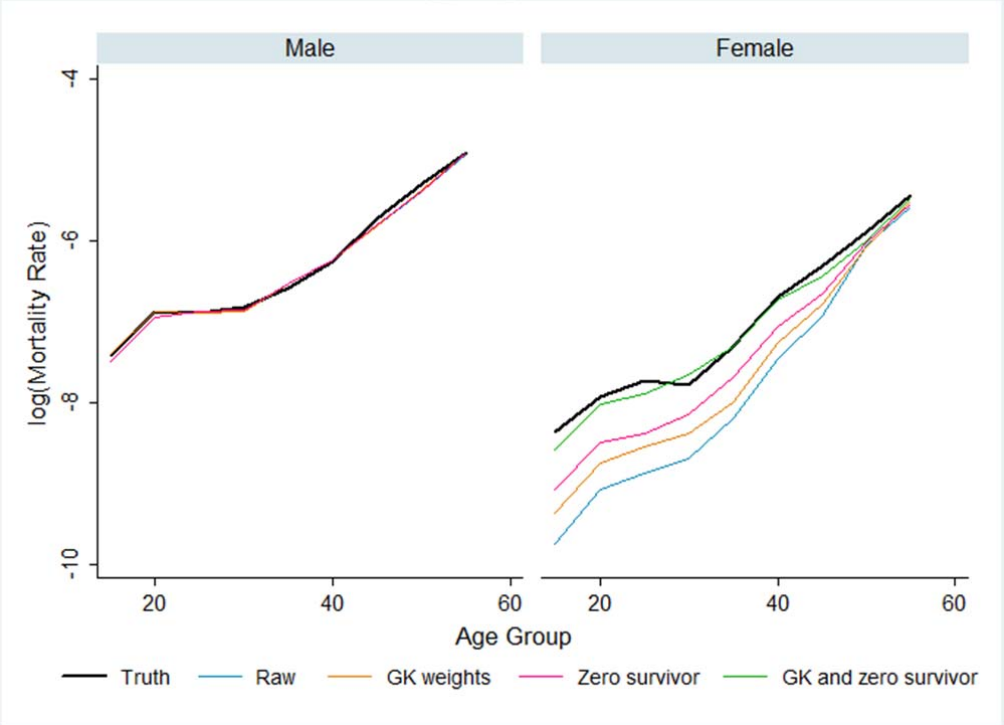

Population Age Distribution by Sex
Female
Male
Age
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
300,000
200,000
100,000
0
100,000
200,000
300,000
Number in Age Group

Figure 2

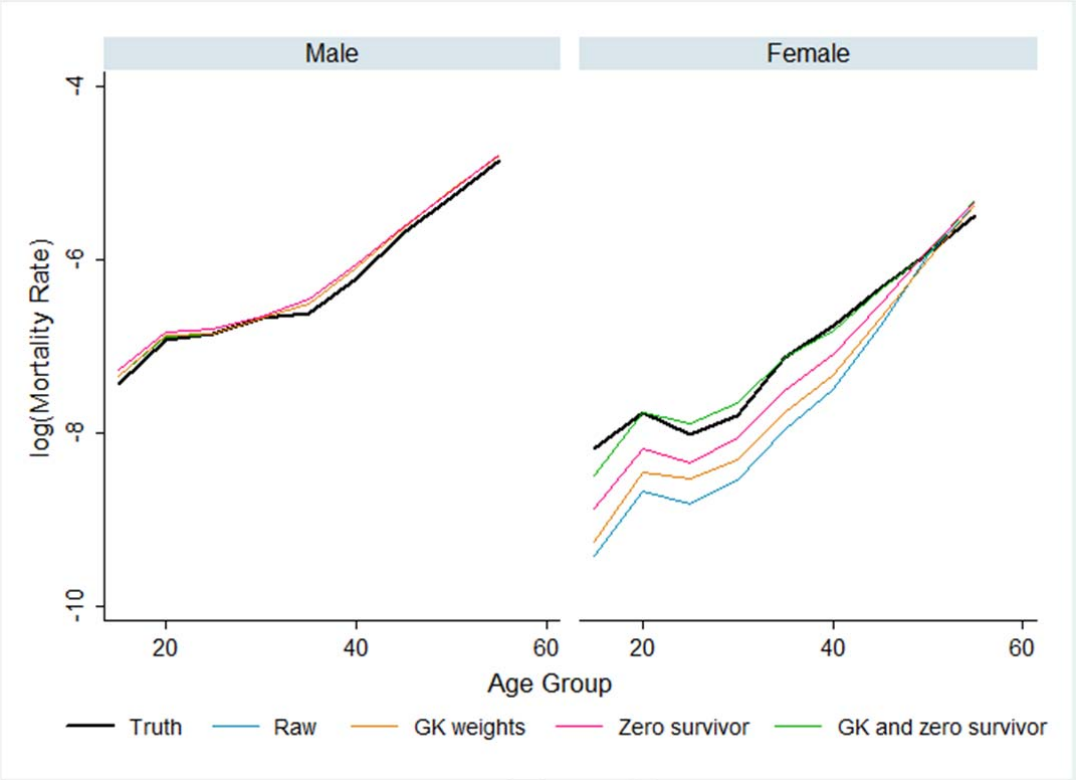
(a)


Male
Female
log(Mortality Rate)
-5
-6
-7
-8
-9
20
40
60
Age Group
Truth
Raw
GK weights
Zero survivor
GK and zero survivor

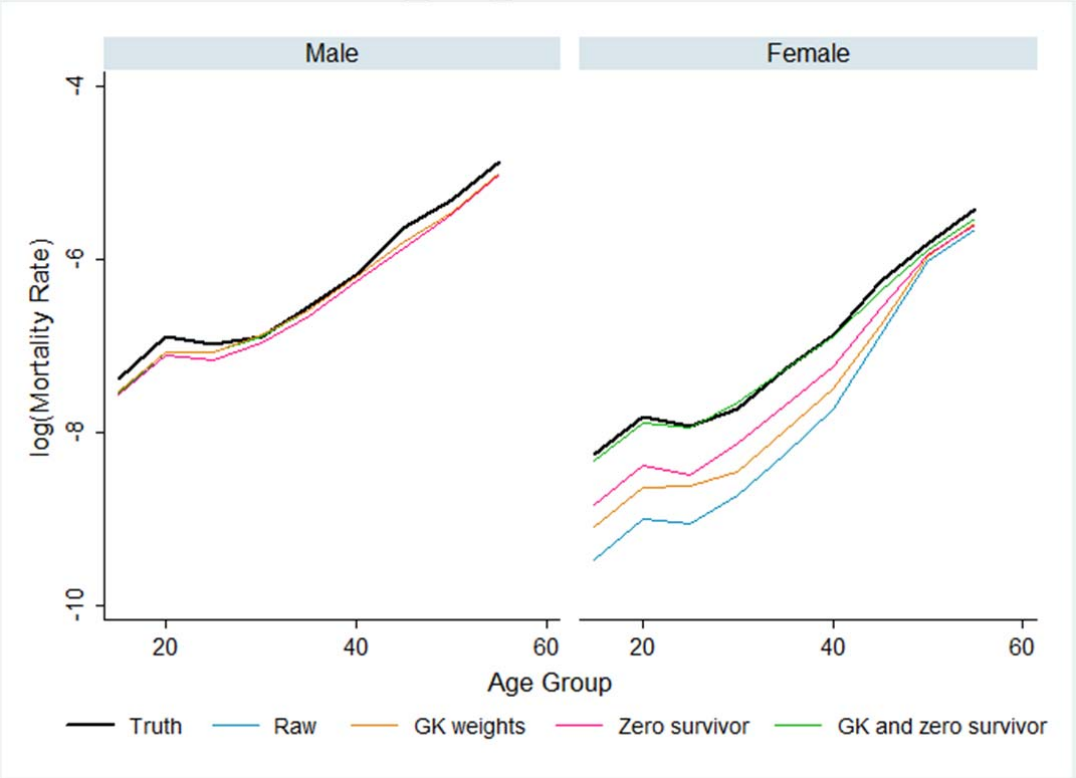
(b)


Male
Female
log(Mortality Rate)
-4
-6
-8
-10
20
40
60
Age Group
Truth
Raw
GK weights
Zero survivor
GK and zero survivor

(c)


Male
Female
log(Mortality Rate)
-4
-6
-8
-10
20
40
60
Age Group
Truth
Raw
GK weights
Zero survivor
GK and zero survivor

(d)


Male
Female
log(Mortality Rate)
-4
-6
-8
-10
20
40
60
Age Group
Truth
Raw
GK weights
Zero survivor
GK and zero survivor

(e)

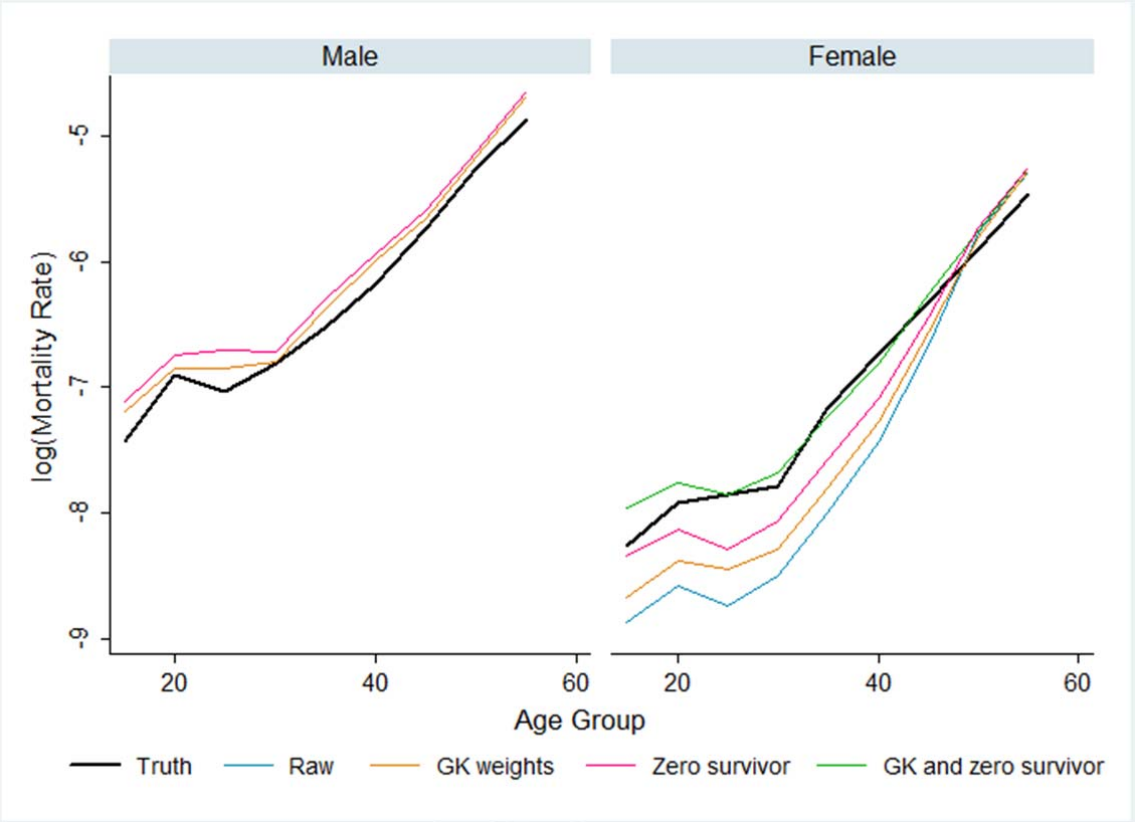

Male
Female
log(Mortality Rate)
-5
-6
-7
-8
-9
20
40
60
Age Group
Truth
Raw
GK weights
Zero survivor
GK and zero survivor

Figure 3

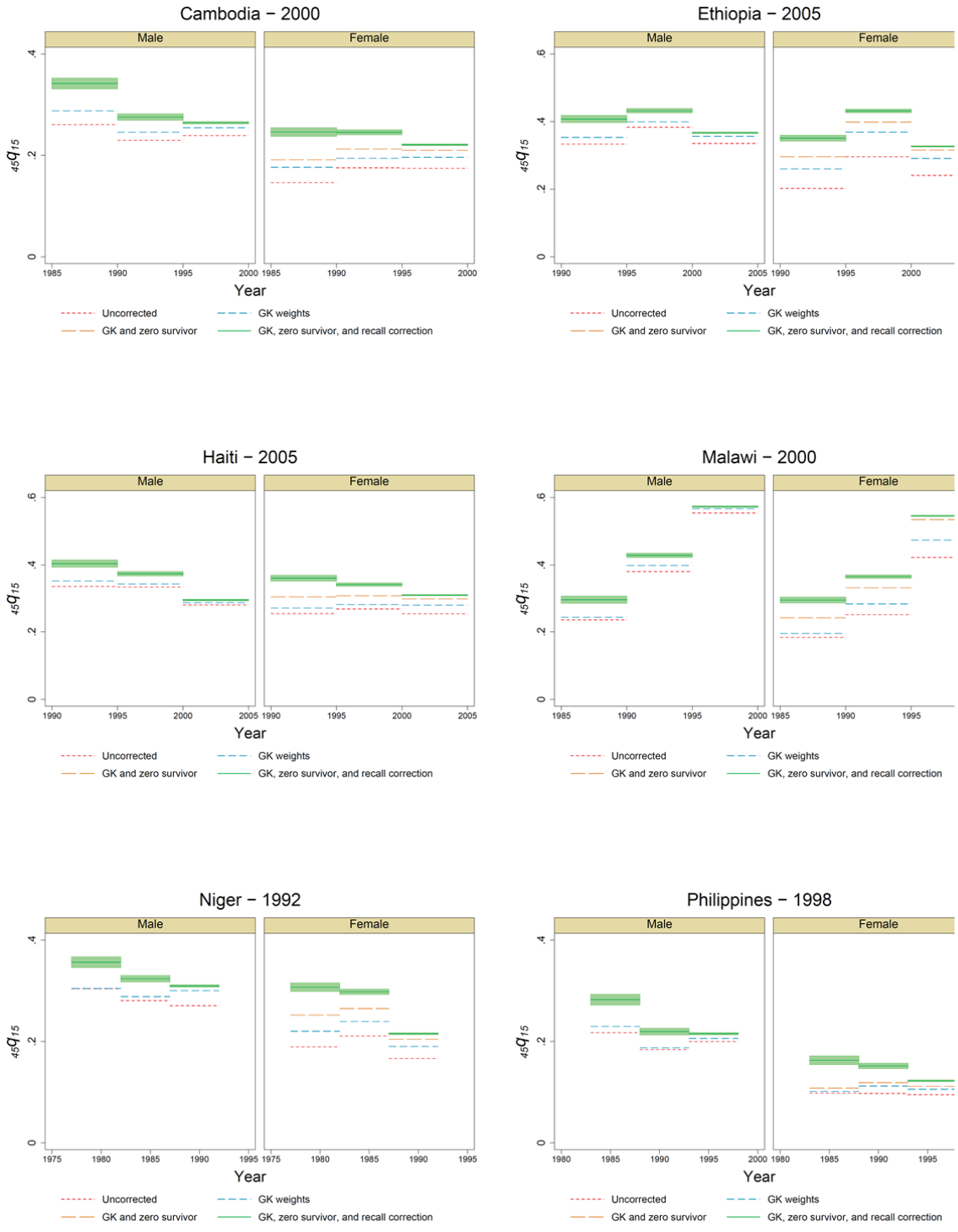

Cambodia − 2000
Male
Female
$_{45}q_{15}$
Year
Uncorrected
GK weights
GK and zero survivor
GK, zero survivor, and recall correction
Ethiopia − 2005
Haiti − 2005
Malawi − 2000
Niger − 1992
Philippines − 1998

Figure 4

Appendices:

Appendix Table 1

Demographic and Health Surveys
Number of countries: 49
Number of surveys: 95
Respondents: Women, age 15-49 (seven surveys also ask a smaller sample of males about their siblings)
Nationally representative

Country	Year	Respondents	Siblings reported	Sibling deaths reported
Benin	1996	5,488	36,367	7,904
Benin	2006	17,358	114,112	18,279
Bolivia	1993	8,603	51,703	7,842
Bolivia	2003	17,251	112,001	17,541
Bolivia	2008	16,939	103,618	14,687
Brazil	1996	2,949	87,522	11,519
Brazil	1996	12,577	87,522	11,519
Burkina Faso	1998	6,427	40,885	7,817
Burkina Faso	2003	12,230	78,494	12,908
Cambodia	2000	15,351	95,593	16,292
Cambodia	2005	16,516	106,595	18,708
Cameroon	1998	5,490	38,420	6,895
Cameroon	2004	10,656	76,151	14,403
Central African Republic	1994	5,884	36,402	5,814
Chad	1996	7,450	48,547	9,727

Chad	2004	6,085	41,932	8,835
Country	Year	Respondents	Siblings reported	Sibling deaths reported
Congo	2005	7,051	47,843	7,665
Congo, Dem. Rep.	2007	9,472	67,876	12,530
Côte d'Ivoire	1994	8,099	52,487	7,852
Côte d'Ivoire	2005	4,891	33,103	5,936
Dominican Republic	2002	11,384	74,643	6,814
Dominican Republic	2007	27,162	163,193	14,336
Eritrea	1995	1,114	32,724	6,594
Eritrea	1995	5,054	32,724	6,594
Ethiopia	2000	15,347	107,524	25,669
Ethiopia	2005	13,739	94,664	17,661
Gabon	2000	6,183	42,410	6,020
Ghana	2007	10,370	63,728	7,193
Guatemala	1995	12,375	84,202	11,990
Guinea	1999	6,753	39,198	7,684
Guinea	2005	7,619	48,978	11,038
Haiti	2000	10,159	68,682	14,601
Haiti	2005	10,523	71,827	14,541
Indonesia	1994	28,168	162,221	22,526
Indonesia	1997	28,810	157,744	15,948
Indonesia	2002	28,051	157,540	14,081
Indonesia	2007	32,895	176,733	17,115
Jordan	1997	5,546	51,077	5,836
Kenya	1998	7,872	57,405	6,192
Kenya	2003	8,177	60,066	7,698

Kenya	2008	8,444	58,248	5,897
Lesotho	2004	6,902	40,632	6,633
Country	Year	Respondents	Siblings reported	Sibling deaths reported
Lesotho	2009	7,246	40,312	6,796
Liberia	2006	6,658	37,887	3,901
Madagascar	1992	6,260	46,494	6,920
Madagascar	1997	7,060	51,649	6,028
Madagascar	2003	7,630	51,165	3,989
Madagascar	2008	17,375	121,092	14,114
Malawi	1992	1,151	35,018	9,190
Malawi	1992	4,849	35,018	9,190
Malawi	2000	13,220	92,513	22,280
Malawi	2004	11,290	74,080	14,047
Mali	1995	9,704	63,045	13,897
Mali	2001	12,815	83,365	16,140
Mali	2006	14,118	100,149	22,193
Mauritania	2000	7,728	49,563	6,242
Morocco	1992	9,256	69,757	10,665
Morocco	2003	16,602	124,923	18,033
Mozambique	1997	8,702	50,149	8,493
Mozambique	2003	11,923	74,937	13,291
Namibia	1992	5,421	36,605	4,231
Namibia	2000	6,752	44,527	4,830
Namibia	2006	9,499	60,920	7,622
Nepal	1996	8,429	52,134	11,474
Nepal	2006	10,639	64,475	11,356

Country	Year	Respondents	Siblings reported	Sibling deaths reported
Niger	1992	6,503	43,819	10,285
Niger	2006	8,935	64,183	13,526
Nigeria	2008	33,385	213,304	30,322
Peru	1991	15,882	99,447	10,480
Peru	1996	28,951	197,378	28,873
Peru	2000	27,774	183,986	25,290
Peru	2003	11,441	73,794	8,564
Peru	2004	40,552	255,728	33,811
Philippines	1993	15,029	102,938	8,132
Philippines	1998	13,978	93,976	8,479
Rwanda	2000	10,415	75,804	19,432
Rwanda	2005	11,184	83,711	22,590
Sao Tome and Principe	2008	2,615	19,326	2,533
Senegal	1992	6,310	41,913	8,160
Senegal	2005	14,370	100,964	15,295
Sierra Leone	2008	6,612	38,903	6,137
South Africa	1998	11,718	63,008	6,551
Sudan	1989	5,860	42,570	6,855
Swaziland	2006	4,810	30,489	3,984
Tanzania	1996	8,118	55,931	8,539
Tanzania	2004	10,190	74,273	13,175
Tanzania	2009	10,139	66,974	8,947
Timor-Leste	2009	13,137	76,631	10,147
Togo	1998	8,569	58,193	11,110
Uganda	1995	1,996	51,196	9,621

Uganda	1995	7,068	51,196	9,621
Uganda	2000	7,240	54,871	11,046
Uganda	2006	8,518	66,655	15,165
Zambia	1996	8,021	59,231	10,075
Country	Year	Respondents	Siblings reported	Sibling deaths reported
Zambia	2001	7,658	55,477	9,859
Zambia	2007	7,141	48,382	7,695
Zimbabwe	1994	2,141	44,649	5,780
Zimbabwe	1994	6,128	44,649	5,780
Zimbabwe	1999	5,907	40,401	4,560
Zimbabwe	2005	7,175	55,810	6,897
Zimbabwe	2005	8,661	55,810	6,897

CDC Reproductive Health Surveys

Number of countries with sibling history survey: 5

Number of surveys: 8

Respondents: Women, age 15-49

Nationally representative

Country	Year	Respondents	Siblings reported	Sibling deaths reported
Ecuador	1994	12,352	76,444	6,639
Ecuador	2004	9,521	50,098	508
El Salvador	1993	5,417	29,713	1,154
El Salvador	1998	11,164	58,893	1,128
El Salvador	2003	9,211	47,975	578

Honduras	1996	6,953	44,625	3,000
Nicaragua	1992	6,446	61,439	2,323
Paraguay	1995	5,837	34,687	2,441

Population Health Metrics Research Consortium Surveys

Number of sites with sibling history survey: 4

Number of surveys: 6

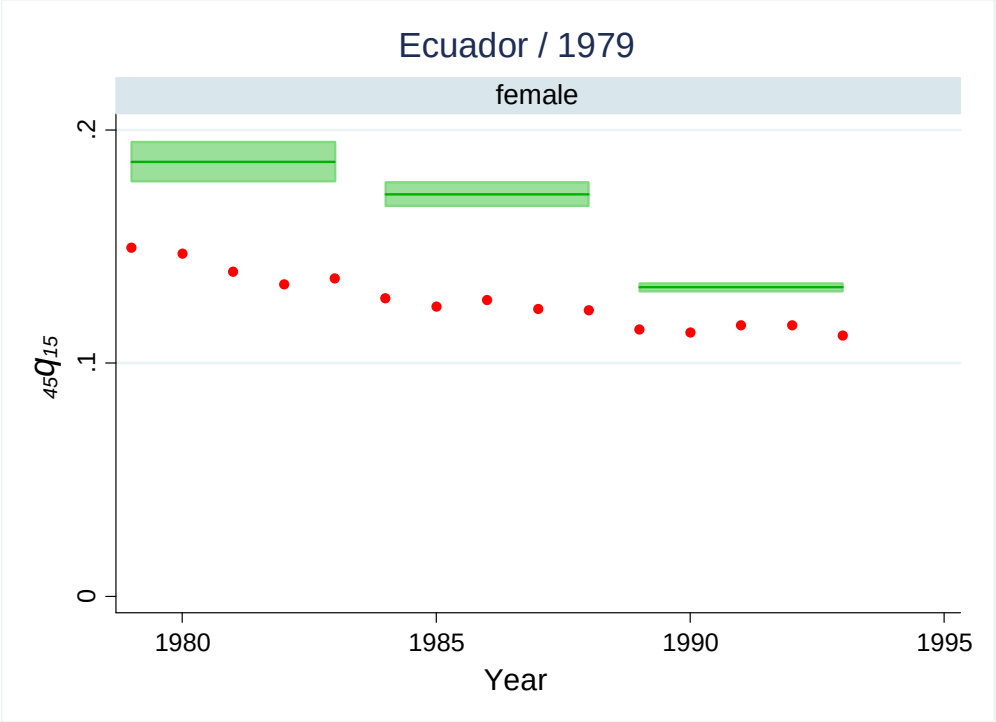
Respondents: Men and women, age 15+

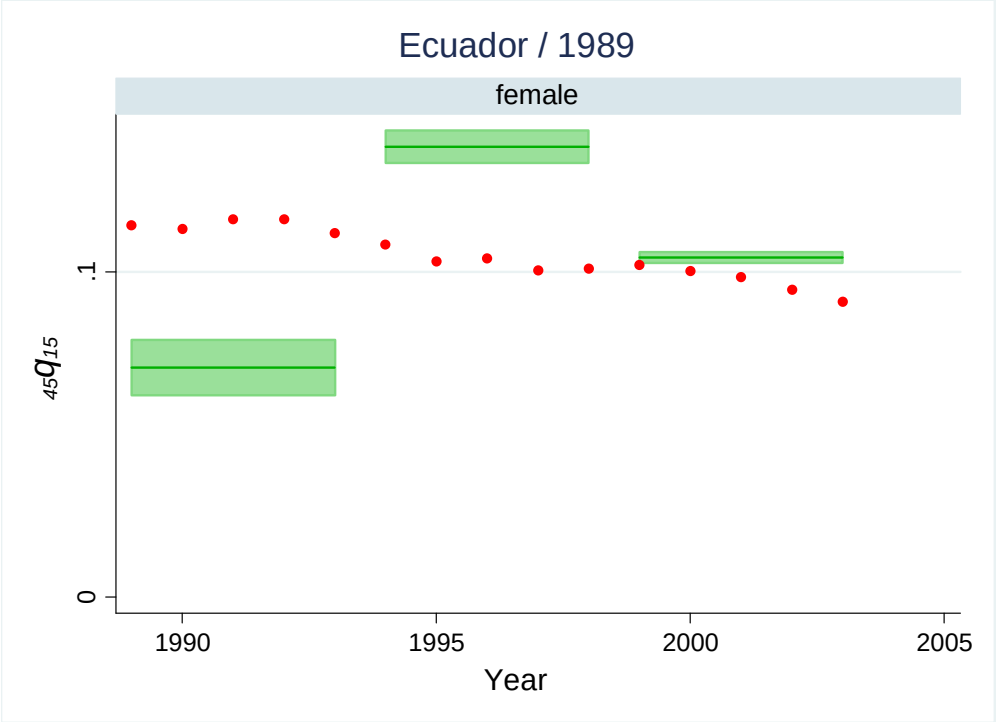
Not nationally representative

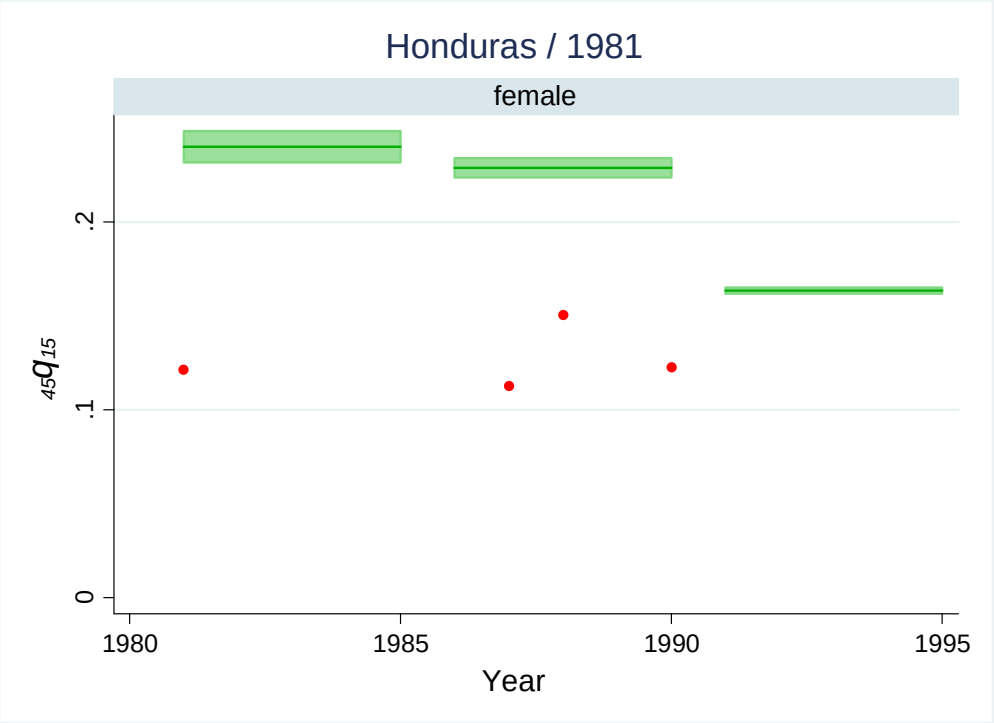
Country	Year	Respondents	Siblings reported	Sibling deaths reported
Andhra Pradesh, India	2007-2009	72,299	248,215	69,404
Uttar Pradesh, India	2008-2009	28,373	112,281	25,792
Bohol, Philippines	2007-2008	28,370	161,411	28,387
Pemba Island, Tanzania	2007-2008	69,083	420,097	58,347
Andhra Pradesh, India	2010	3,943	14,639	4,801
Bohol, Philippines	2010	5,027	29,271	4,950

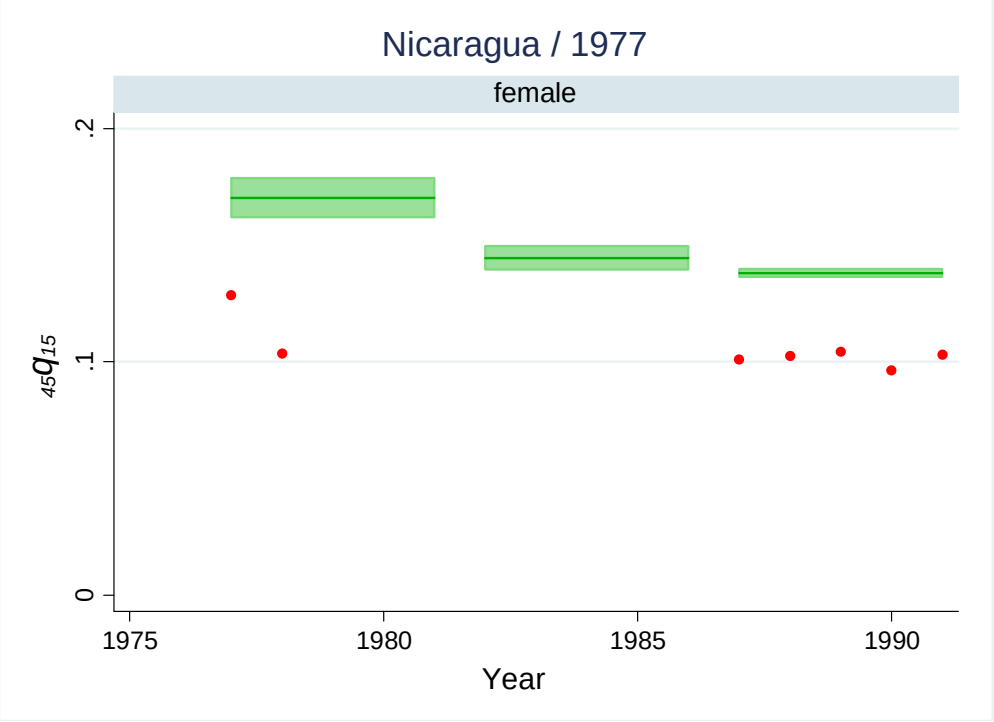
Appendix Figure 1: Estimates of 45q15 from sibling history surveys as compared to estimates from other independent sources of adult mortality, for each country where both sources exist.

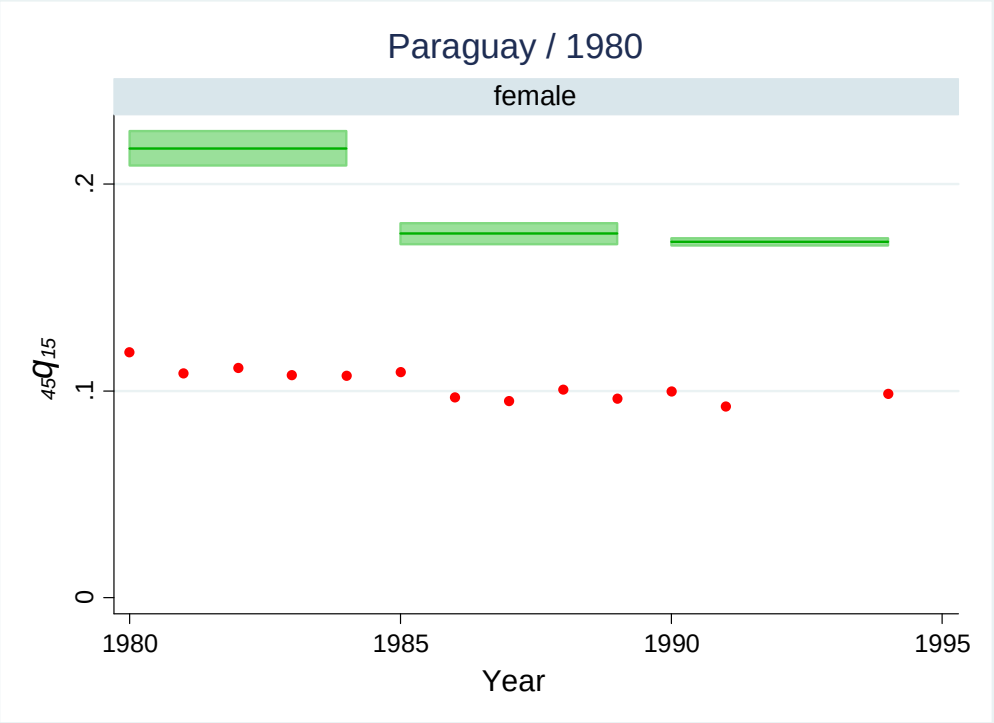
Estimates for Reproductive Health Surveys:

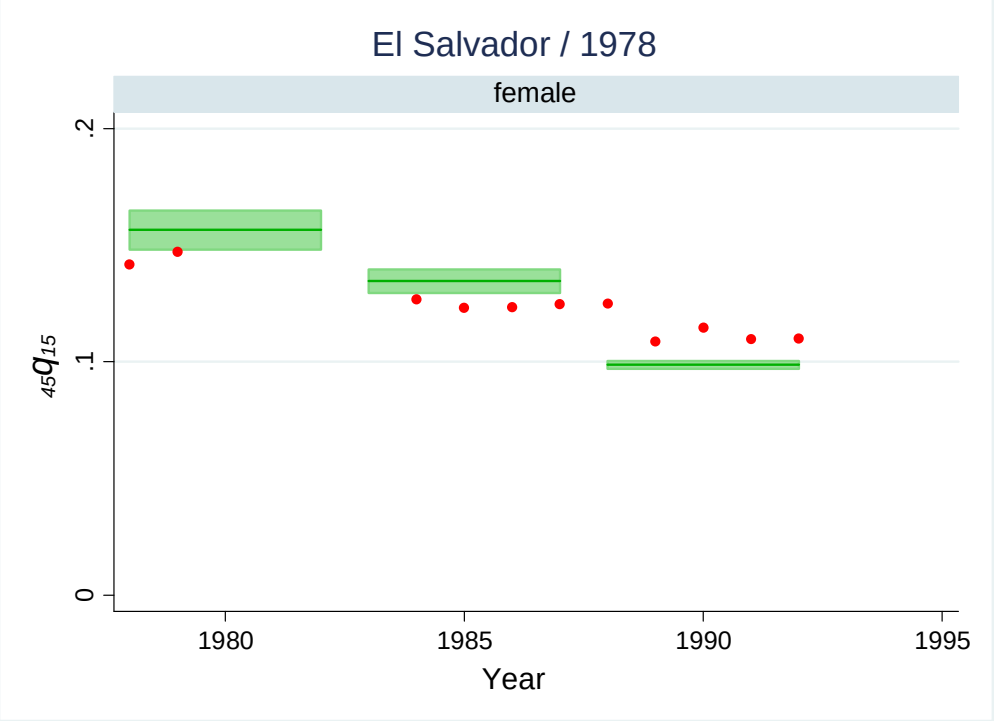


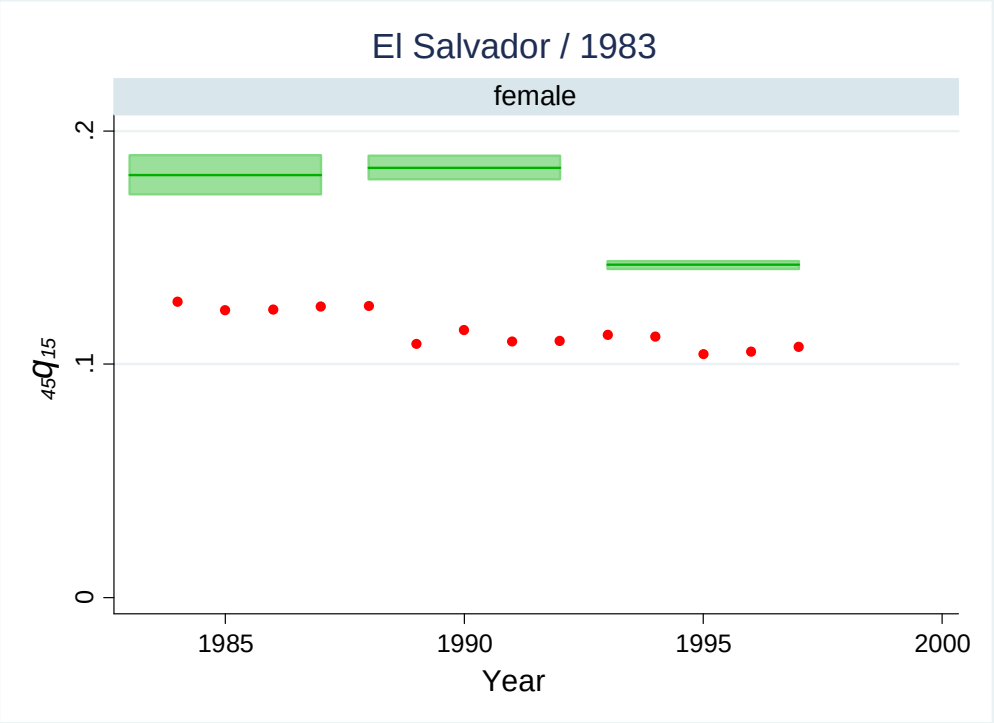


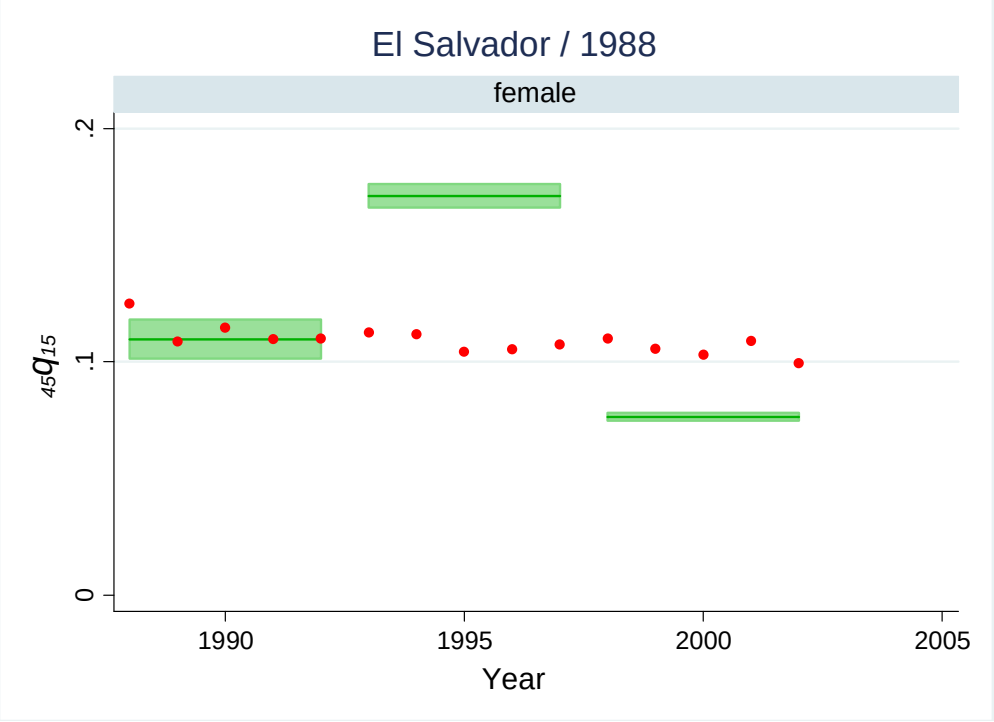

Honduras / 1981
female
$_{45}q_{15}$
0
.1
.2
1980
1985
1990
1995
Year


Nicaragua / 1977
female
$_{45}q_{15}$
0
.1
.2
1975
1980
1985
1990
Year

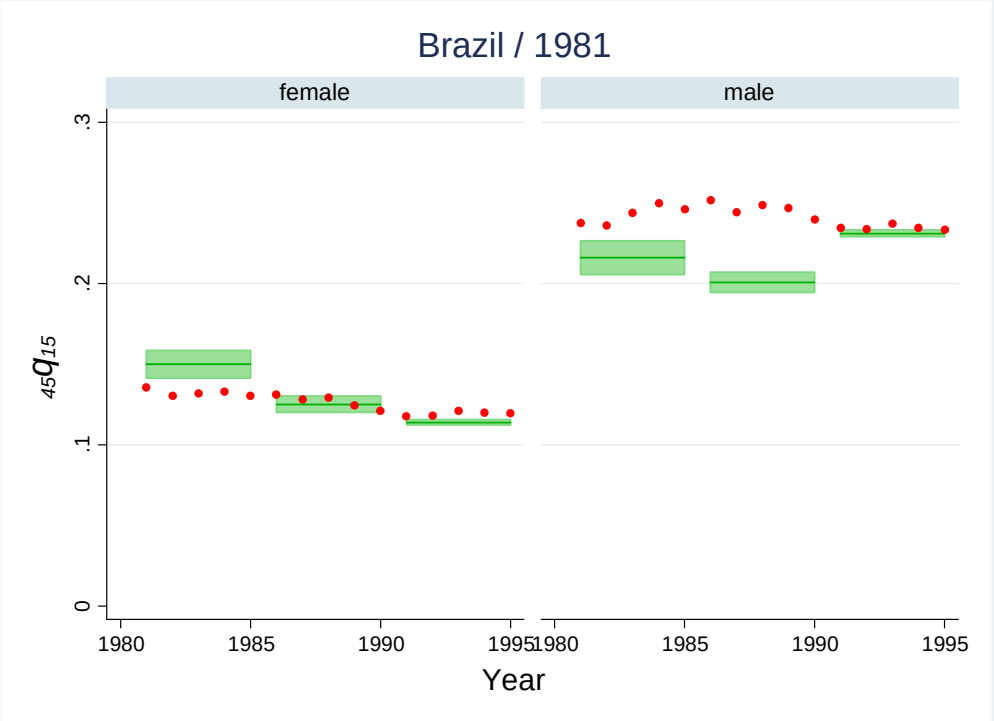

Paraguay / 1980
female
$_{45}q_{15}$
0
.1
.2
1980
1985
1990
1995
Year

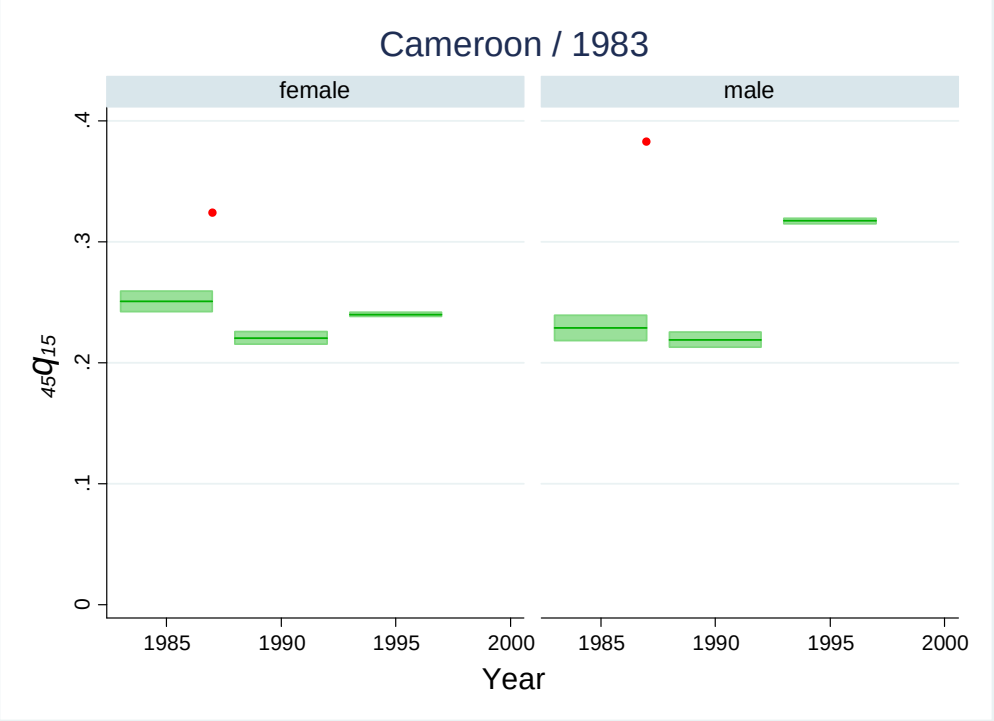

El Salvador / 1978
female
$_{45}q_{15}$
0
.1
.2
1980
1985
1990
1995
Year

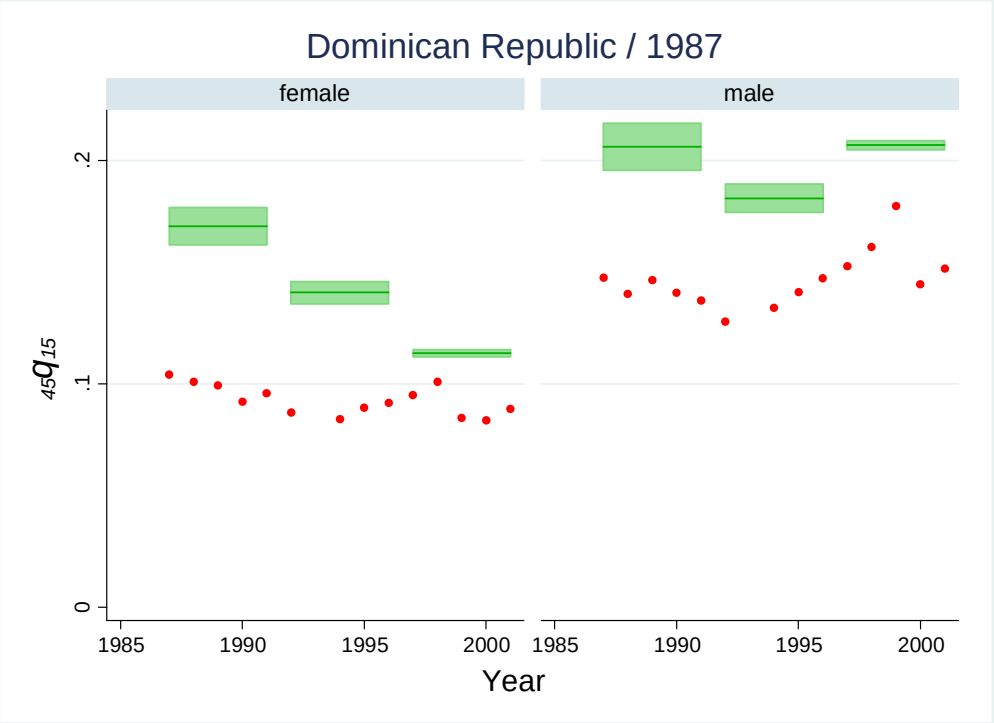

El Salvador / 1983
female
.2
.1
0
45q15
1985
1990
1995
2000
Year

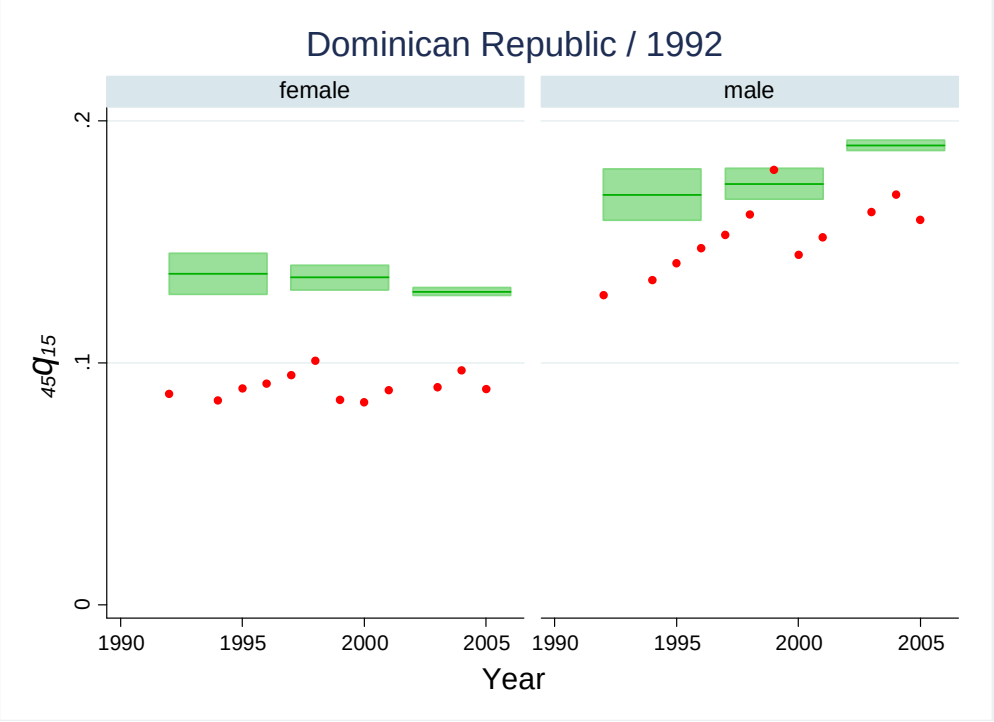

El Salvador / 1988
female
.2
.1
0
45q15
1990
1995
2000
2005
Year

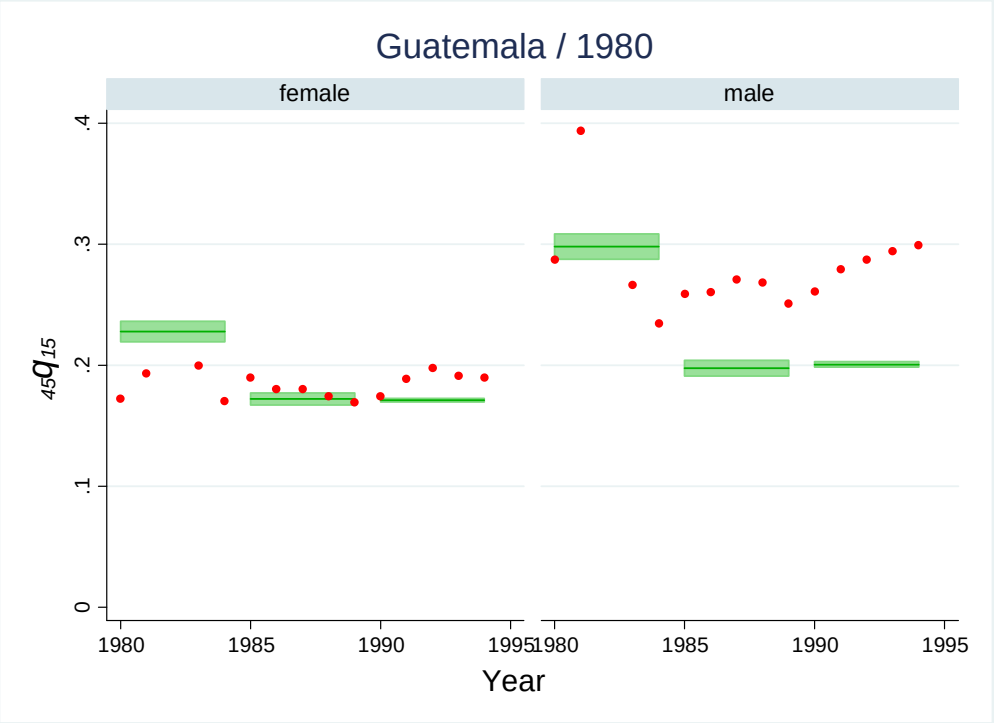
Estimates for Demographic Health Surveys:

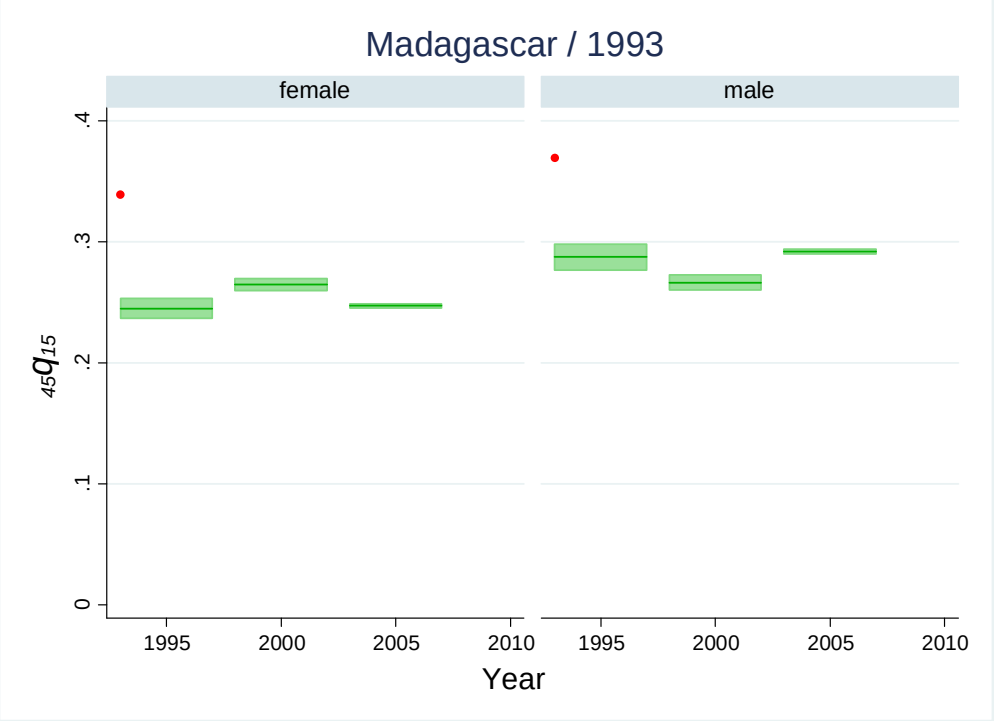


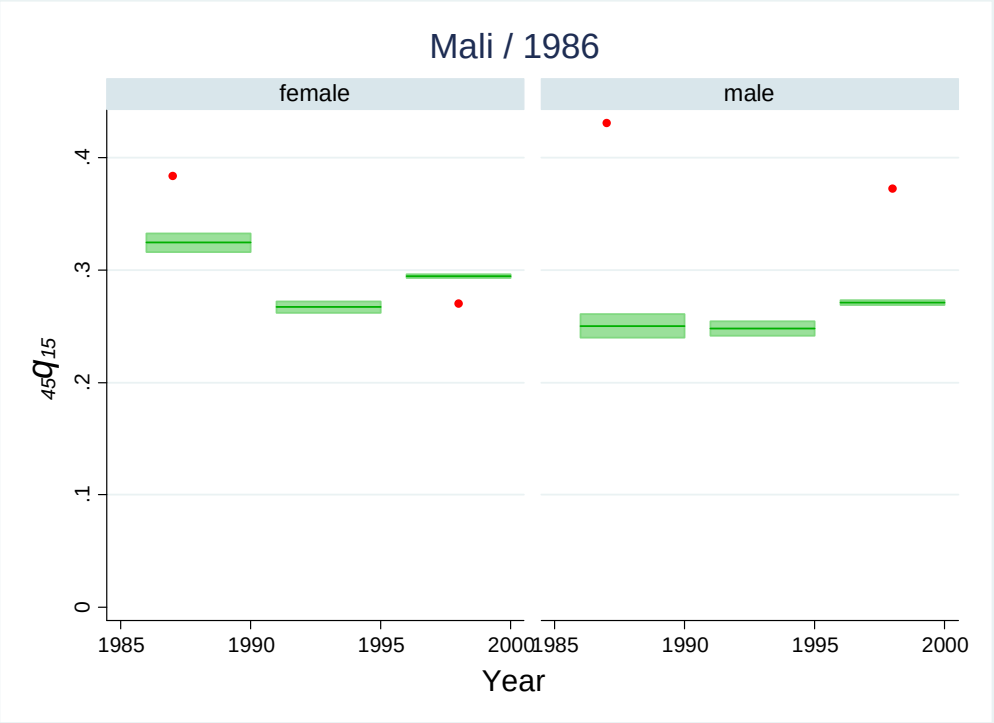


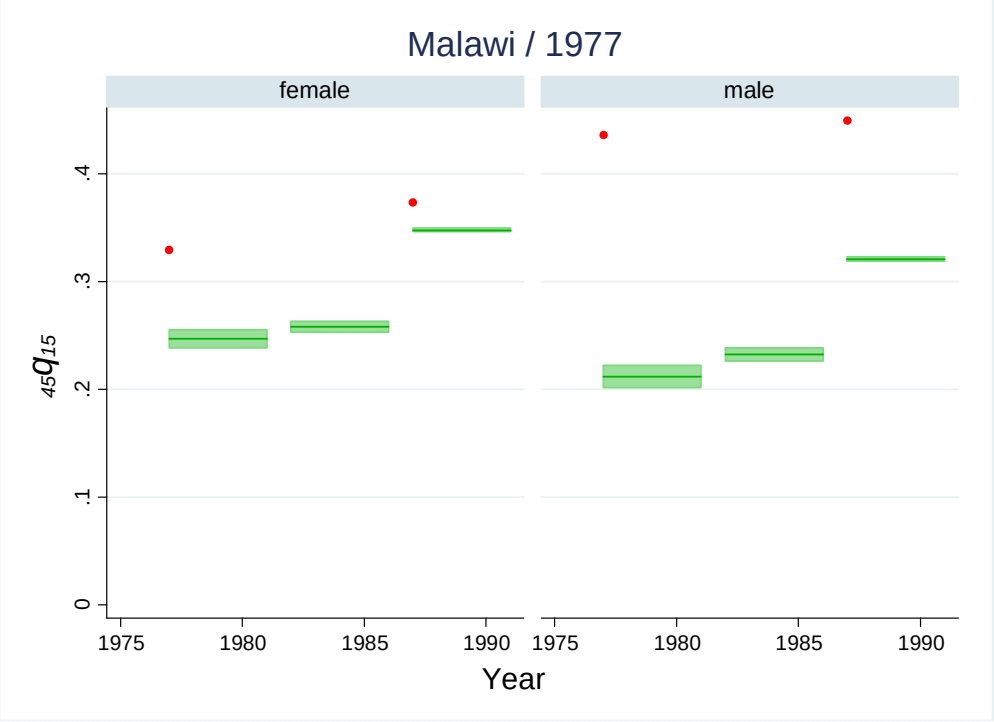

Dominican Republic / 1987
female
male
$_{45}q_{15}$
0
.1
.2
1985
1990
1995
2000
Year

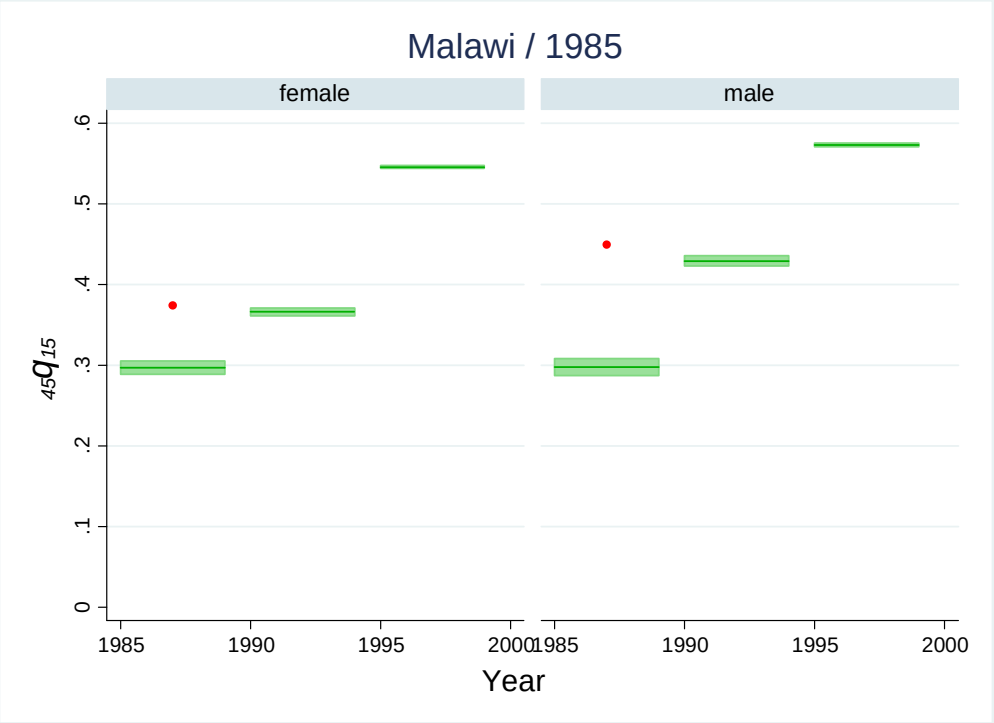

Dominican Republic / 1992
female
male
$_{45}q_{15}$
0
.1
.2
1990
1995
2000
2005
Year

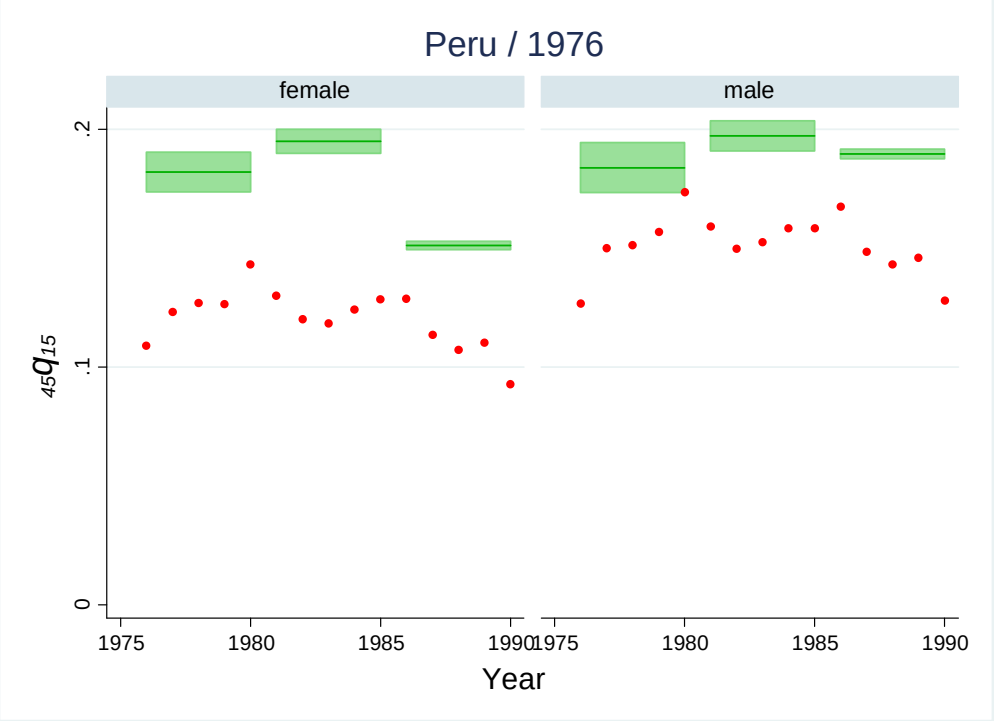

Guatemala / 1980
female
male
$_{45}q_{15}$
0
.1
.2
.3
.4
1980
1985
1990
1995
Year

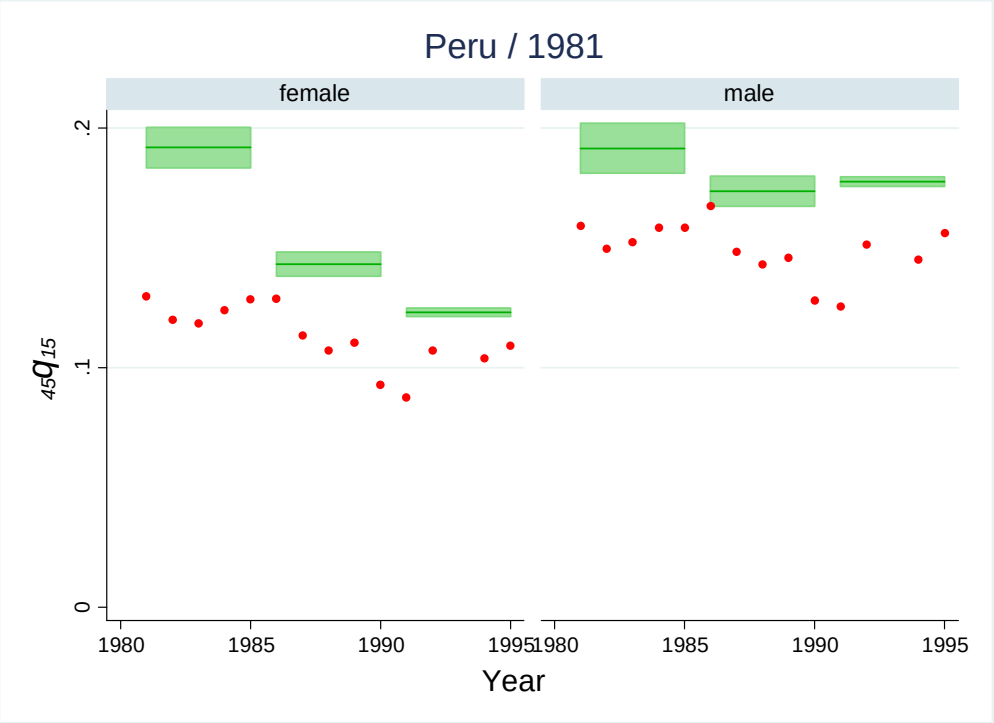

Madagascar / 1993
female
male
$_{45}q_{15}$
0
.1
.2
.3
.4
1995
2000
2005
2010
Year

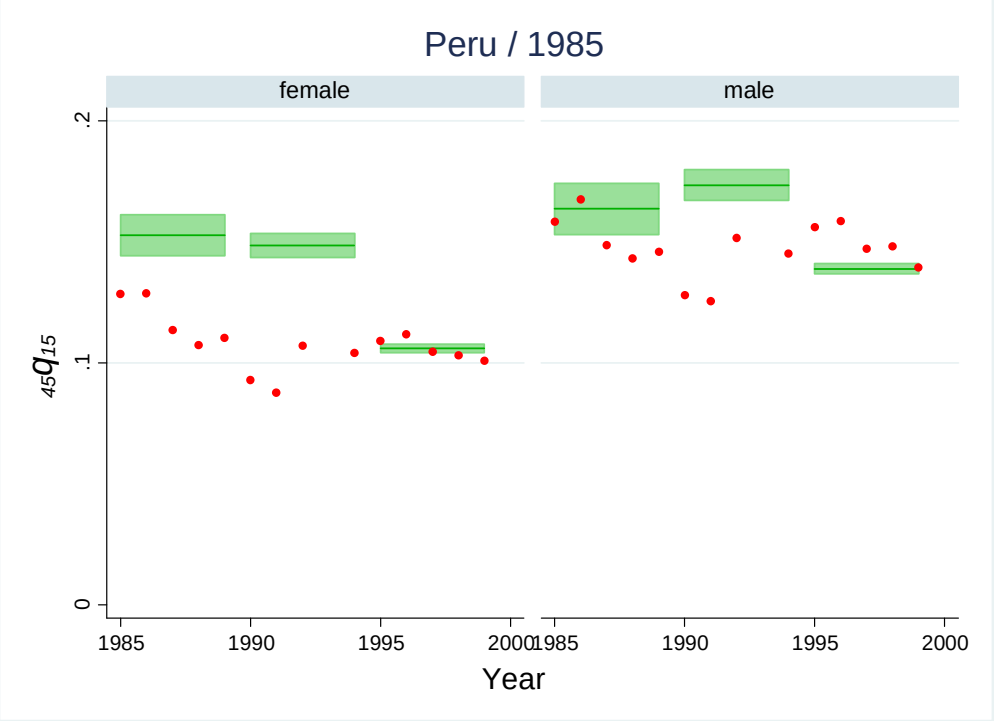

Mali / 1986
female
male
$_{45}q_{15}$
0
.1
.2
.3
.4
1985
1990
1995
2000
Year

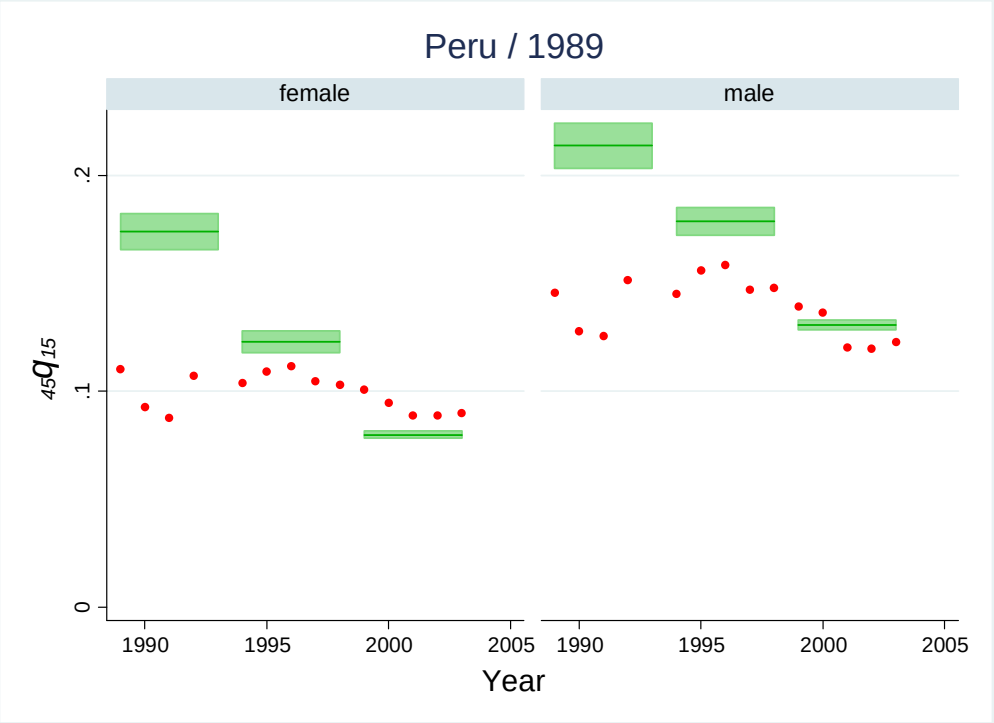

Malawi / 1977
female
male
$_{45}q_{15}$
0
.1
.2
.3
.4
1975
1980
1985
1990
Year

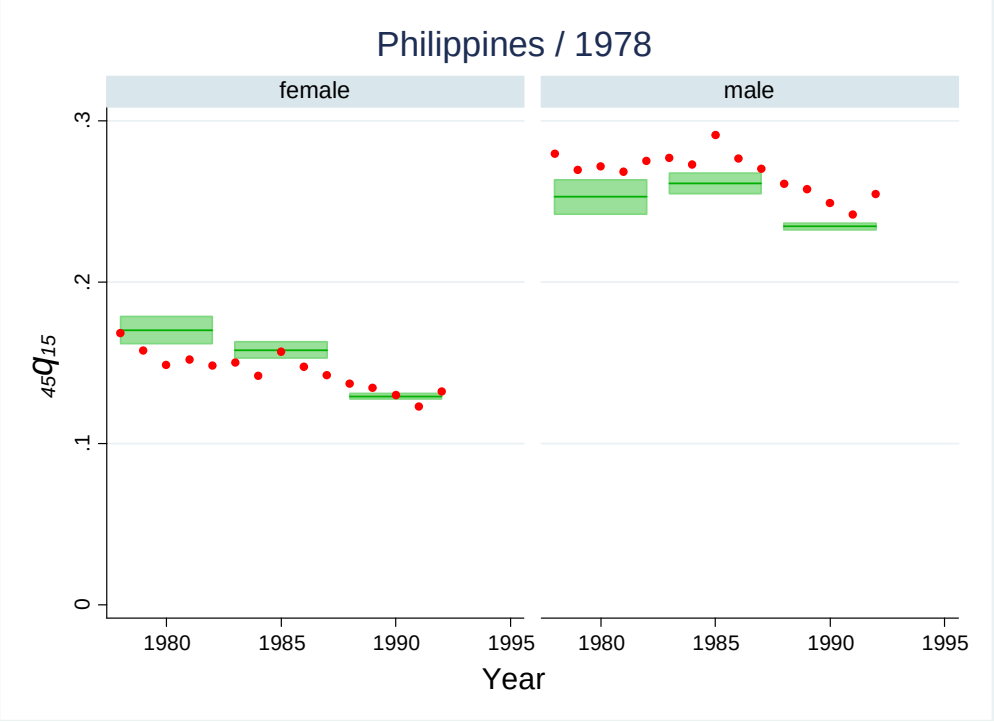

Malawi / 1985
female
male
$_{45}q_{15}$
0
.1
.2
.3
.4
.5
.6
1985
1990
1995
2000
Year

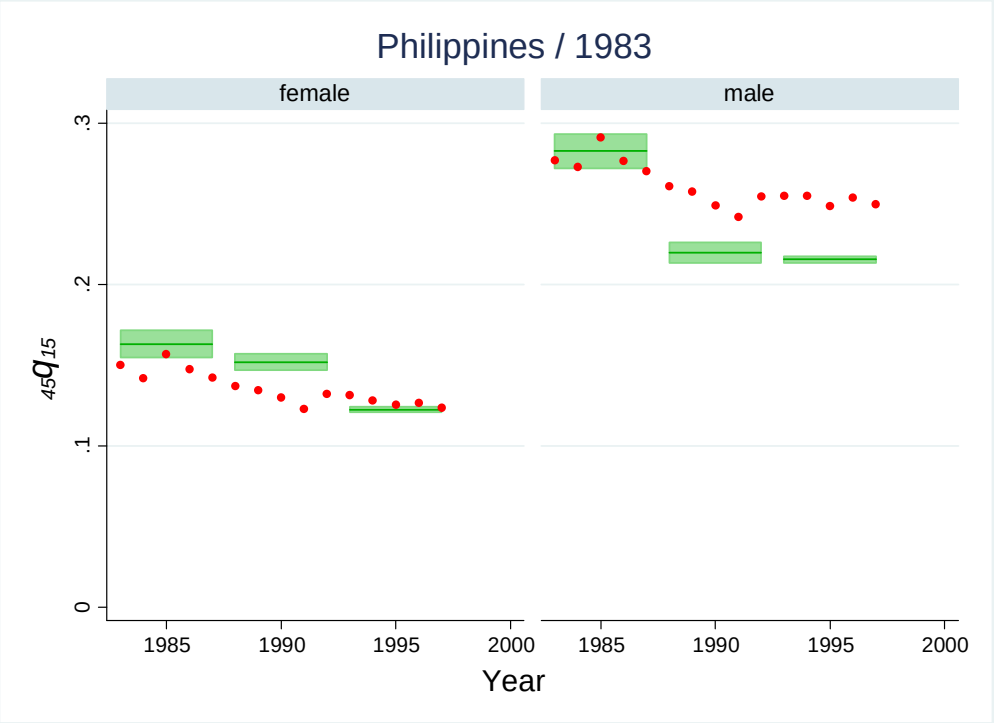

Peru / 1976
female
male
$_{45}q_{15}$
0
.1
.2
1975
1980
1985
1990
Year

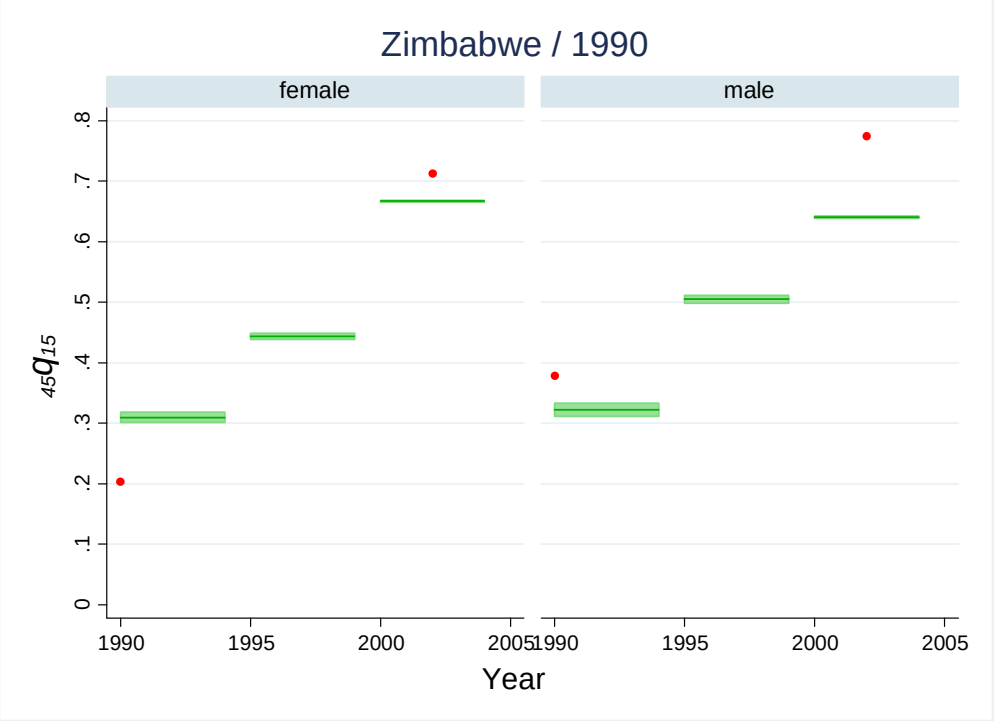

Peru / 1981
female
male
0
.1
.2
45q15
1980
1985
1990
1995
Year


Peru / 1985
female
male
0
.1
.2
45q15
1985
1990
1995
2000
Year


Peru / 1989
female
male
0
.1
.2
45q15
1990
1995
2000
2005
Year


Philippines / 1978
female
male
0
.1
.2
.3
45q15
1980
1985
1990
1995
Year


Philippines / 1983
female
male
$_{45}q_{15}$
0
.1
.2
.3
1985
1990
1995
2000
Year


Zimbabwe / 1990
female
male
$_{45}q_{15}$
0
.1
.2
.3
.4
.5
.6
.7
.8
1990
1995
2000
2005
Year

Appendix

Expanded zero-reporter correction explanation
We use the following equations to directly estimate the number of missing single-sister sibships due to zero-reporter bias within each 5-year age group.

$$K^1_{obs} = K^1_{true} * (1 - {}_aq^1_0) \quad (1)$$

$$K^1_{miss} = K^1_{true} *_a q^1_0 \quad (2)$$

$$\therefore\ K^1_{miss} = \frac{K^1_{obs}}{1 - {}_aq^1_0} *_a q^1_0 \quad (3)$$

In words, equation (1) says that the number of sibships with one sister that are observed in the sampled population, K^1_{obs}, is equal to the true number of sibships with one sister in the population, K^1_{true}, multiplied by the probability that the sister has survived to the time of the survey, where her cumulative probability of death is denoted as aq^1_0 for her 5-year age group a. Equation (2) says that the number of sibships with one sister that are not represented in the sampled population due to zero-reporter bias, K^1_{miss}, is equal to the true number of sibships with one sister in the population, K^1_{true}, multiplied by the probability that the sister has died before the time of the survey. It follows that the number of sibships with one sister that are not represented in the sampled population due to zero-reporter bias is equal to equation (3). We can multiply this estimate of the number of missing sibships by the number of females in the sibship, which in this case is one, to get an estimate of the number of females in each age group that are missing from the sample because they have died. We then expand this number so that we have on observation per missing sibling, assign birth and death dates to these missing siblings based on the distribution in the observed siblings, and append them to our existing dataset.

Our estimate of aq^1_0 is the GK weighted cumulative probability of death for age group a. By definition, this is an underestimate because we are omitting women in zero-reporter sibships from our calculation. To account for this, we treat the estimation of missing siblings as an iterative process and update our estimate of aq^1_0 at the end of each cycle until we converge on a single value of aq^1_0 and missing female siblings.

As noted before, we can expand this estimation procedure to families with two sisters (or two siblings of either sex in the case of the PHMRC data). The corresponding equations when both sisters are within the 15-49 age range are as follows.

$$K^2_{obs} = K^2_{true} * (1 - aq^1_0 * aq^2_0) \quad (4)$$

$$K^2_{miss} = K^2_{true} * aq^1_0 * aq^2_0 \quad (5)$$

$$\therefore K^2_{miss} = \frac{K^2_{obs}}{1-aq^1_0 * aq^2_0} * aq^1_0 * aq^2_0 \quad (6)$$

Where K^2_{obs} is the number of observed sibships with two sisters in our sample, K^2_{true} is the true number of sibships with two sisters in the population, K^2_{miss} is the number of sibships with two sisters that were missed due to zero-reporter bias, aq^1_0 is the cumulative probability of death for the first sister and aq^2_0 is the cumulative probability of death for the second sister. Equation (6) is solved in an iterative manner for every combination of five-year age groups in the sample.

Similarly, for sibships of two sisters in which the first sister is within the 15-49 range but the second is not, the equations are different because the survival of the second sister does not contribute to the probability of the sibship being observed in the sample. The corresponding equations are as follows and the estimation process is conceptually identical to the previous two cases.

$$K^2_{obs} = K^2_{true} * (1 - aq^1_0) \quad (7)$$

$$K^2_{miss} = K^2_{true} * aq^1_0 * aq^2_0 + K^2_{true} * aq^1_0 * (1 - aq^2_0) \quad (8)$$

$$\therefore K^2_{miss} = \frac{K^2_{obs}}{1-aq^1_0} * aq^1_0 * aq^2_0 + \frac{K^2_{obs}}{1-aq^1_0} * aq^1_0 * (1 - aq^2_0) \quad (9)$$