An Assessment of DHS Estimates of Fertility and Under-Five Mortality

Thomas W. Pullum[1]

Bruno Schoumaker[2]

Stan Becker[3]

Sarah E.K. Bradley[4]

August 2013

Prepared for the 2013 International Population Conference of the International Union for the Scientific Study of Population (IUSSP), Session 132: Data quality in demographic surveys, August 28. This research was funded by the United States Agency for International Development (USAID), under a contract with ICF International (ICFI). This paper is a partial draft of a Methodological Report that will be published by the Demographic and Health Surveys in September 2013. This paper will be superseded by that report. Much of the emphasis here will be on strategies rather than findings.

[1] Demographic and Health Surveys, ICFI; tom.pullum@icfi.com
[2] Centre de recherche en démographie et sociétés / IACCHOS, Université catholique de Louvain; bruno.schoumaker@uclouvain.be
[3] Bloomberg School of Public Health, Johns Hopkins University; sbecker@jhsph.edu
[4] Department of Demography, University of California at Berkeley, and ICFI; sarah.bradley@icfi.com

Introduction

DHS surveys are the main source of estimates of levels, differentials, and trends in fertility and under-five mortality in developing countries. Births recorded in the retrospective birth histories permit direct estimation of age-specific fertility rates and the Total Fertility Rate. Information about the survivorship of children and when they died (if they died) also permit direct estimation of neonatal, post-neonatal, infant, child, and under-five mortality rates. When combined with relatively recent statistical approaches, the analysis of such data can include confidence intervals, hypothesis tests, multi-level modeling, projections, and other sophisticated approaches to the analysis of fertility and mortality. All of these measures and analyses are based on the underlying accuracy of the events and dates in the birth histories. The main concerns of the proposed paper are with (a) evidence of weaknesses in the birth histories and (b) the sensitivity of standard fertility and mortality rates to such weaknesses.

This report has three main sections. In Part 1, two potential types of errors in the data will be considered. The first is possible omission of births in the birth history. Such omissions are known to be more likely for children who died very young, particularly as neonatal deaths. Such omissions may also be more likely for specific intervals of time before the interview, for specific types of women, and for specific interviewers. Omissions of children who died will induce a downward bias in both fertility rates and mortality rates. A second potential weakness is systematic displacement of events. There is known to be some displacement of births backwards and out of the window for the numerous questions on the health of children. Other kinds of displacement are associated with specific ages and dates and are manifested as heaping of responses on those ages and dates. For example, a child's completed age at death is sometimes stated as seven days, one month, or twelve months, when the death actually occurred within the first week or first month or first year, respectively.

Part 2 of the report will investigate the hypothesis that the quality of the birth histories has been negatively affected by the increasing complexity of the surveys. It would be reasonable to expect a progression toward better data quality in more recent surveys, because of improvements in the expertise of the implementing agencies, the increasing educational levels of interviewers and respondents, better technology for data collection, etc. However, it has been suggested that DHS surveys have tended to deteriorate in quality because of a trend toward greater overall complexity of the surveys. The gradual addition of more questions, modules, and biomarkers has led to longer interviews, and potentially to interviewer and respondent fatigue, and a possible degradation of the quality of the birth histories.

Part 3 of the report will apply an innovative adjustment procedure to correct the birth histories for possible omission and displacement. By comparing the adjusted and unadjusted rates we can assess the sensitivity of the estimates of the published fertility and mortality rates to such errors, regardless of how the errors may have arisen during the data collection. The deviations in rates that appear to be induced by data quality issues will be compared with the estimated standard errors of those rates. We will be particularly concerned with deviations that exceed two standard errors, implying statistical significance at approximately the .05 level.

Part 1. Evidence of omission and displacement in the birth histories [5]

1.1. Consistency of fertility estimates across surveys

The first approach consists in reconstructing fertility trends by calendar year over the last fifteen years in all the surveys. The method, using Poisson regression (Schoumaker, 2013a), allows reconstructing trends of the TFR between 15 and 49 and for about 15 years, and evaluating the consistency of estimates across surveys. The visual inspection of the reconstructed trends suggests that data quality varies greatly across surveys and countries. It also illustrates different types of data quality problems. Six situations are described below.

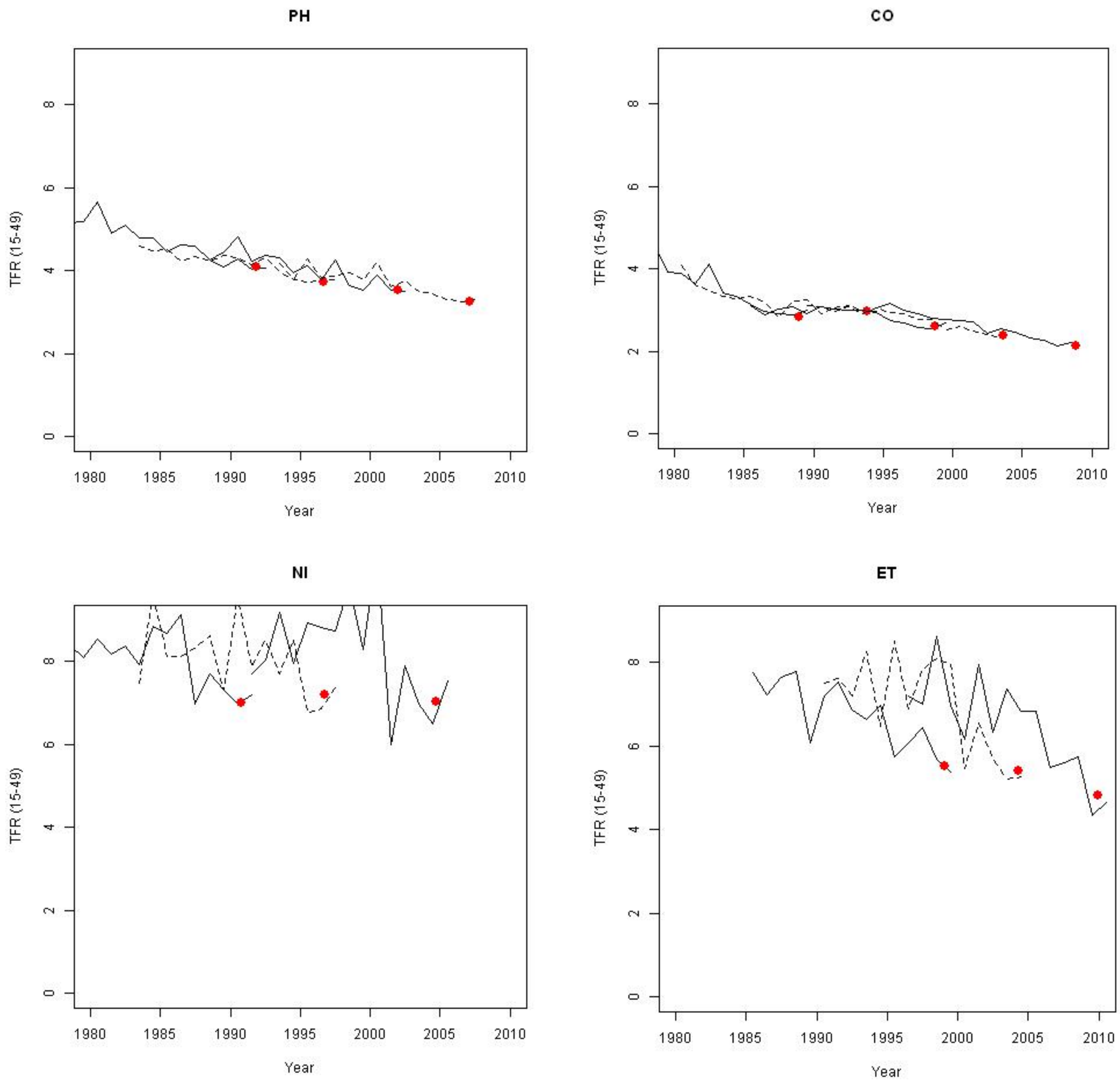
The first figure shows fertility trends from four surveys in the Philippines. Fertility trends in the Philippines show a high degree of consistency across surveys. Annual variations in the TFR are small, the reconstructed trends match quite well, and the published TFRs (last 3 years) are located along the trend. A similar situation is found with the five surveys in Colombia. Even though the retrospective estimates do not match perfectly, they are close to one another. At the opposite, fertility estimates in Niger seem to be much less reliable. Annual values of the TFRs vary widely, and recent estimates are much lower than TFR estimated at the same time in the following surveys. The apparent underestimation of recent fertility may be the consequence of displacements and/or omissions of births. Ethiopia also seems to be affected by severe data quality problems. Published TFRs are well below the other estimates, also suggesting possible displacements and/or omissions. Intermediate situations are illustrated by data from Ghana and Bolivia. In Ghana, TFRs fluctuate to a larger extent than in the Philippines, and the published TFR in the 1998 survey (second survey shown) is below the trend. In Bolivia, fluctuations are also larger than in the Philippines and Colombia, and two estimates of recent fertility appear to be below the general trend. In addition to possible displacements and/or omissions of births, it seems the estimate from the second survey may be lower than the other estimates because of difference in sample implementation.

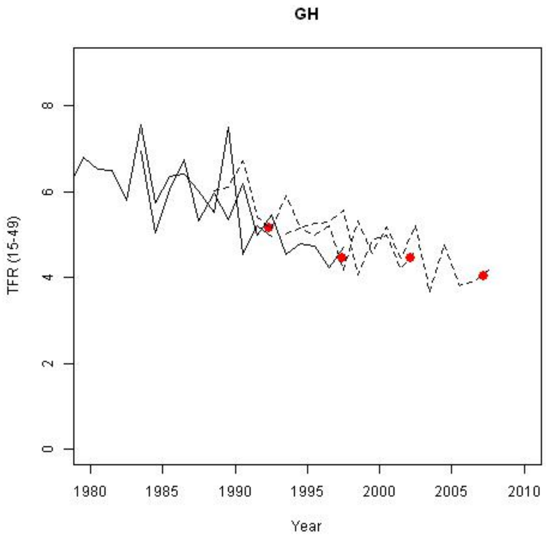
At this stage it is worth mentioning that some fluctuations may reflect real fluctuations in the TFR, and some fluctuations reflect sampling errors. However, given that sample sizes are fairly similar across surveys (check), the greater fluctuations probably partly reflect data quality problems.

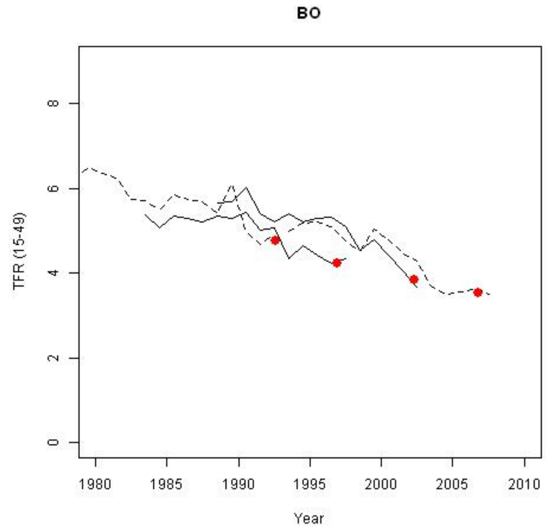
Visual inspection suggests data quality problems are more severe in sub-Saharan African countries, as well as in some specific countries outside Africa (e.g. Yemen, Pakistan, Haiti, Bangladesh).

[5] Bruno Schoumaker is the principal author of this section of the paper.

Figure 1.1 : Comparison of retrospective estimates of fertility in successive surveys and published TFRs (Philippines, Colombia, Niger, Ethiopia, Ghana, Bolivia).







1.2. Reconstruction of fertility by pooling birth histories

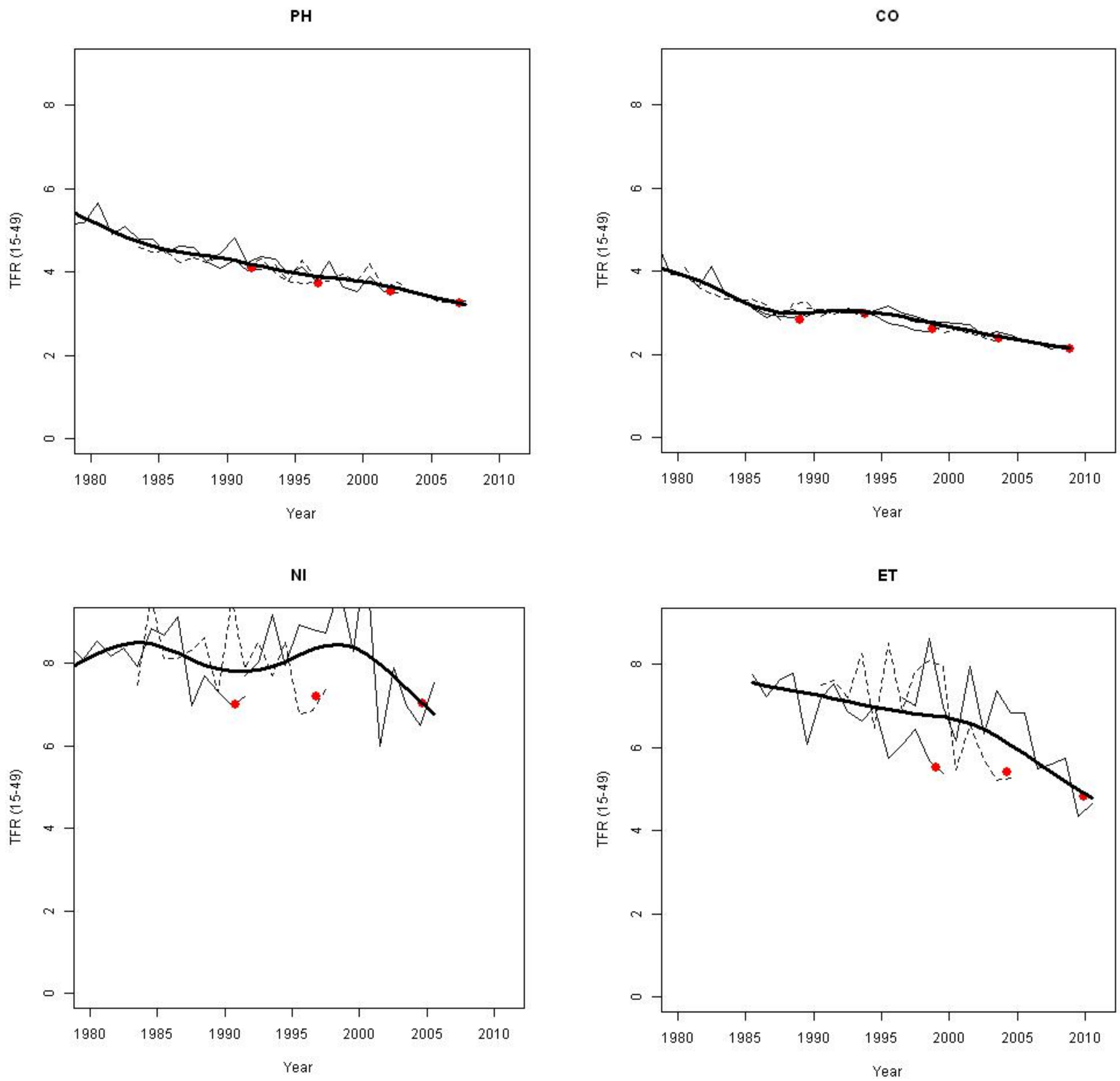
Pooling data from successive surveys allows reconstructing smoothed fertility trends over a relatively long period. The method, described in Schoumaker (2013b), consists in pooling tables of births and exposure from all the surveys (for fifteen years before each surveys), and reconstructing the trends with Poisson regression and restricted cubic splines. Varying age schedules of fertility are allowed, and simulations show the method performs very well in reconstructing long term fertility trends with good quality data (Schoumaker, 2013b).

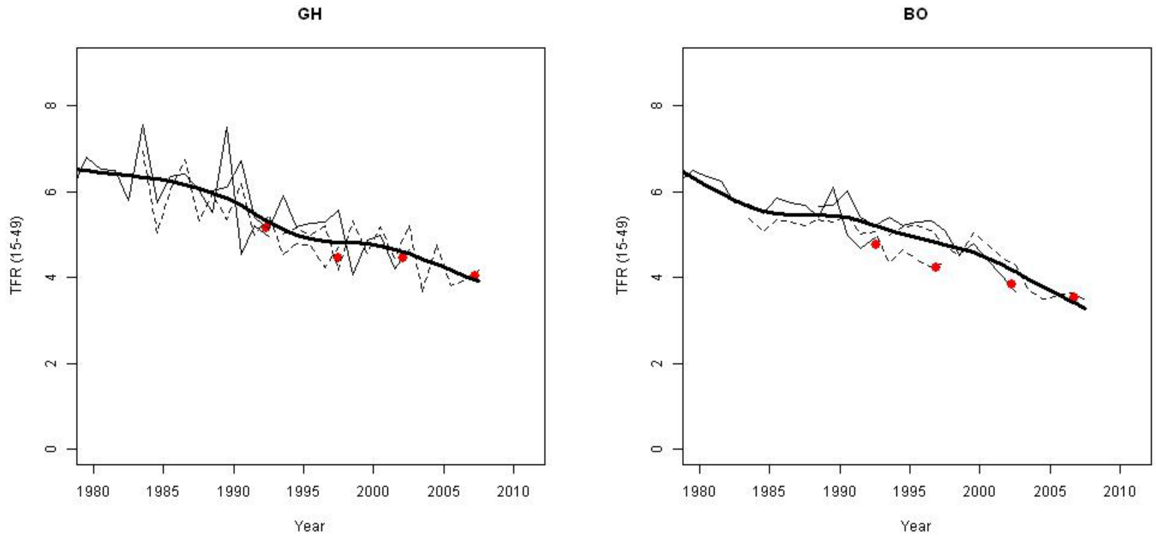
The method is applied in all the countries. As illustrated on Figure 1.2, published TFRs are very close to the reconstructed fertility trend in the Philippines and Colombia. In contrast, discrepancies are much higher in Niger and Ethiopia. In both countries, TFRs fluctuate widely and published TFRs are much lower than the TFR estimated by pooling the surveys for the first two surveys[6]. For the most recent survey, the reconstructed estimate and the published estimate are obviously very close to each other. The Ghana and Bolivia cases illustrate less severe data quality problems. In both countries, however, the published TFR is lower than the reconstructed TFR.

The reconstruction of fertility in all the countries shows that – overall – consistency across surveys is overall quite good. Most countries with severe inconsistencies appear to be in sub-Saharan Africa.

[6] Where pooled estimates rely on more than one data source.

Figure 1.2 : Reconstruction of fertility trends (TFR 15-49) and comparison with published TFRs (Philippines, Colombia, Niger, Ethiopia, Ghana, Bolivia).



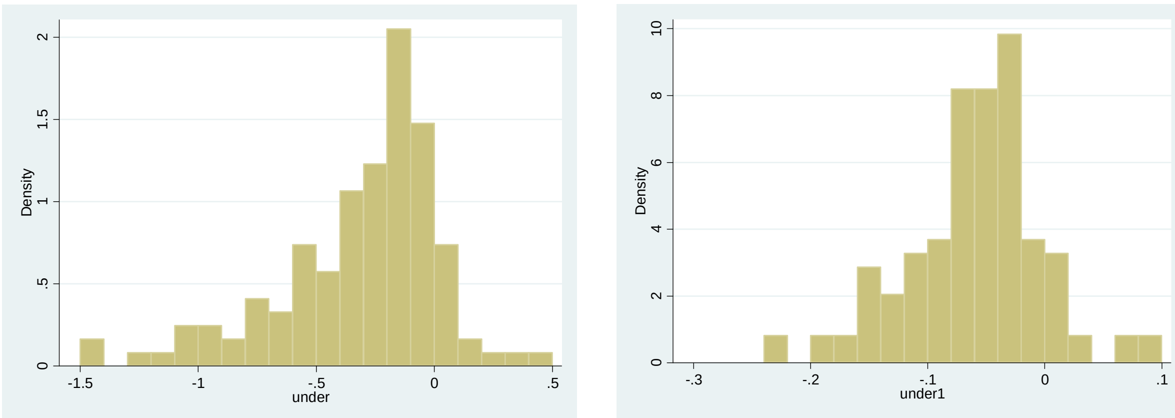


1.3. Underestimation of fertility

The reconstructed estimates can be used to evaluate – in a rough way– the extent to which published fertility estimates (last 3 years) are underestimated. This is done in the following way: for all the surveys except the most recent one in each country, the published estimate (red dot) is compared to the reconstructed estimate (black line). As shown on figure 1.2, when the consistency across surveys is good, the published estimate is close to the reconstructed estimate. In contrast, the published TFR can be much lower than the estimate from the reconstructed trend. Although the reconstructed estimate cannot be considered to be the true estimate, we expect it to be closer to the real value of the TFR. The reconstructed estimate at one point in time is a (weighted) average of fertility estimated from different surveys. The case of Ethiopia clearly illustrates that the black thick line (reconstructed fertility) falls between the low estimate from a survey (recent estimate) and high estimates (retrospective estimates obtained from following surveys). The discrepancy between estimates from several surveys may be due to the underestimation of recent fertility, and/or to the overestimation of fertility at the same period in the following survey. The approach adopted here considers that both underestimation and overestimation are at play, and that the reconstructed value of the TFR is a reasonable estimate of the real fertility level.

Figure 1.3 : Distribution of absolute and relative differences between published estimates of TFR (15-49 years, last 3 years) and estimates from reconstruction of fertility trends (15-49 years, centered on the same date) (122 DHS).

(a) Absolute differences (b) Relative differences



The difference between the published and reconstructed TFR is computed for 122 surveys. Countries with only on survey are by definition excluded, and the last survey for all the countries is also excluded. Both the absolute difference and the relative difference between the published and the reconstructed estimates are computed[7].

[7] The difference is compute as ABS=TFR(p)-TFR(r), and the relative difference is REL= TFR(p)/TFR(r)-1.

Figure 1.3(a) and 1.3(b) show the distribution of absolute and relative differences in all the surveys. Overall, differences are moderate in most countries. The large majority of surveys have absolute differences between 0 and 0.5 children, and relative differences between 0 et -0.10. The mode of absolute differences is at -0.2 children, and relative differences peak at -4%. However, some surveys are characterized by large differences. The surveys with the largest relative or absolute differences shown in Table 1.1.

Table 1.2 shows the mean values of this indicator by region, by period and by DHS phase. On average, published TFRs are around 6% lower (0.3 children) than the reconstructed TFRs. The relative differences do not vary a lot across regions, but absolute differences are highest in sub-Saharan Africa and lowest in MENA countries. Overall, differences increased in the third phase of DHS (late 1990s), but have decreased since then. The greater differences between published estimates and reconstructed estimates in the phase III reflect the trend in sub-Saharan Africa. It seems the surveys conducted in the late 1990s in sub-Saharan Africa were affected by greater underestimation of fertility. As argued elsewhere (Schoumaker, 2010), this may be related to the fact that a short reference period (< 4 years) was used for the health module in many surveys in sub-Saharan surveys in the late 1990s. We come back to this issue later.

Table 1.1: Surveys with the largest relative differences between published and reconstructed TFR (<-10 %).

Country	Year	Survey	Relative difference	Absolute difference
Bangladesh	1994	BDIR31	-0.18	-0.78
Bangladesh	1997	BDIR3A	-0.14	-0.53
Bangladesh	2000	BDIR41	-0.10	-0.37
Benin	2001	BJIR41	-0.10	-0.65
Bolivia	1998	BOIR3B	-0.12	-0.58
Burkina Faso	2003	BFIR43	-0.14	-0.99
Burkina Faso	1993	BFIR21	-0.13	-1.00
Cameroon	1998	CMIR31	-0.14	-0.82
Cameroon	2004	CMIR44	-0.12	-0.65
Chad	1997	TDIR31	-0.14	-1.04
Dominican Republic	1999	DRIR41	-0.17	-0.54
Ethiopia	2005	ETIR51	-0.11	-0.70
Ethiopia	2000	ETIR41	-0.18	-1.23
Guinea	1999	GNIR41	-0.16	-1.05
Guyana	2005	GYIR51	-0.10	-0.29
Haiti	1994	HTIR31	-0.14	-0.78
India	1999	IAIR42	-0.17	-0.58
Mali	1996	MLIR32	-0.13	-0.95
Mozambique	1997	MZIR31	-0.15	-0.94
Niger	1992	NIIR22	-0.10	-0.81
Niger	1998	NIIR31	-0.14	-1.15
Nigeria	1999	NGIR41	-0.24	-1.48
Nigeria	2003	NGIR4B	-0.11	-0.70
Pakistan	1991	PKIR21	-0.23	-1.49
Peru	1992	PEIR21	-0.13	-0.52
Turkey	1993	TRIR31	-0.11	-0.32

Table 1.2: Mean of relative and absolute (estimated) underestimation of fertility by region and DHS phase

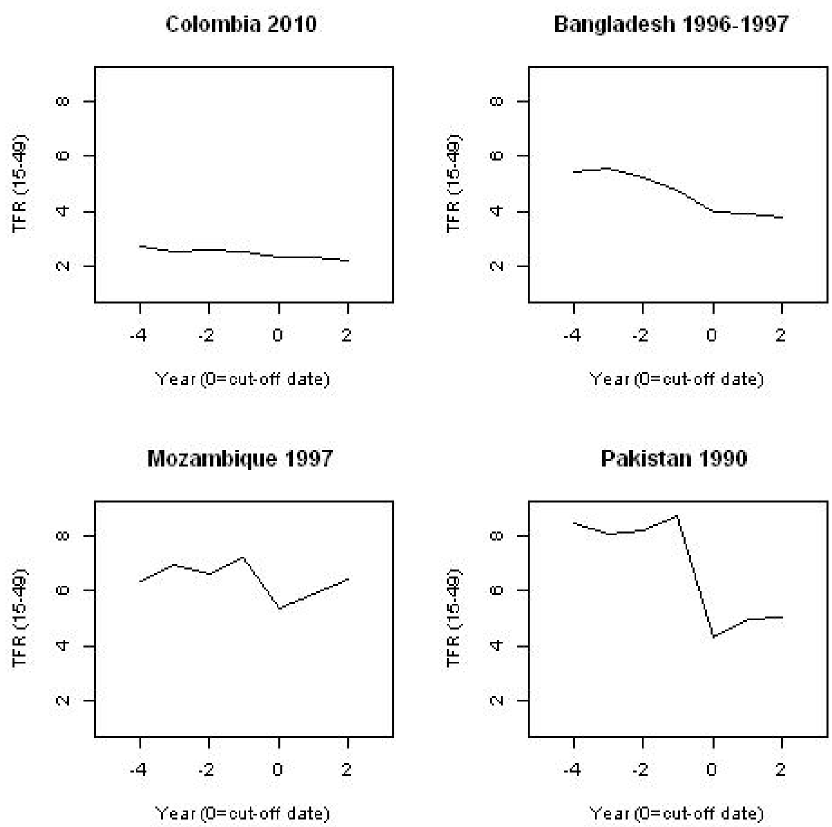
	Relative			Absolute			
Region	Mean	Min	Max	Mean	Min	Max	N
Sub-Saharan Africa	-6.2%	-23.9%	8.9%	-0.40	-1.49	0.48	66
Latin America	-6.2%	-16.9%	1.4%	-0.25	-0.78	0.04	21
MENA	-3.9%	-11.5%	1.1%	-0.14	-0.32	0.04	12
Asia	-5.8%	-23.3%	9.8%	-0.24	-1.49	0.15	23
DHS Phase							
II	-5.1%	-23.2%	7.4%	-0.29	-1.49	0.37	25
III	-8.6%	-23.9%	1.1%	-0.45	-1.49	0.06	42
IV	-5.1%	-18.2%	9.8%	-0.28	-1.23	0.25	40
V	-2.2%	-7.6%	8.9%	-0.12	-0.43	0.48	15
VI							
Total	5.9%	-23.9%	9.8%				122

1.4. Patterns of fertility around the cut-off year of the health module

In order to detect patterns of omissions and/or displacement of births, fertility rates are computed in each survey (except the few surveys with truncated birth histories) by single year for the four years preceding the cut-off date and the three years following the cut-off date of the health module[8]. In the absence of omissions and displacements of births around the cut-off year of the health module, we expect total fertility rates to vary little around the general trend.

The following figures (Figure 1.4) illustrate four cases with varying degrees of distortions in the pattern of TFRs. The first is the most recent DHS in Colombia. Fertility is low and slightly decreasing. The cut-off year of the health module is – apparently – not creating distortions in the pattern of the TFR. In contrast, the 1990 Pakistan survey illustrates a very high degree of distortions. The TFR drops from 8 to 4 children between the year just-before the cut-off year and the cut-off year. This likely reflects displacements and/or omissions of births. The Mozambique case also illustrates a relatively highly distorted pattern of TFR by year, while the Bangladesh case suggests a moderate influence of the cut-off year.

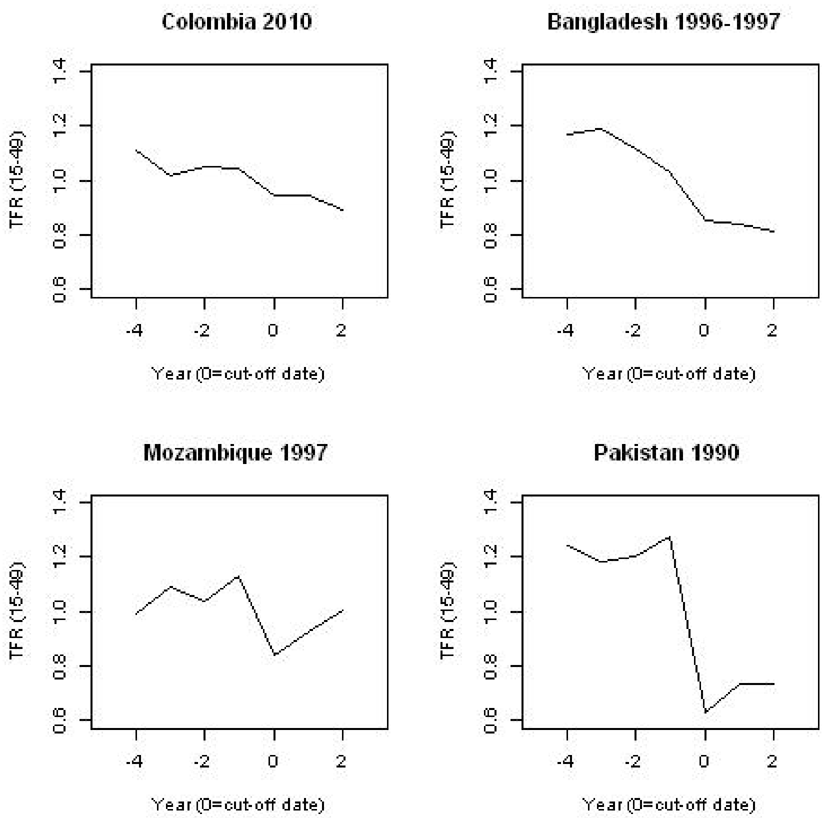
Figure 1.4: patterns of TFRs around the cut-off year of the health module in four DHS



[8] While a four year period after the cut-off date would have been preferred, the cut-off date was three years before the survey in a few cases (e.g. Ghana 1998, Kenya 1998).

Dividing the annual TFR by the average TFR over the seven-year period (relative values of TFRs) facilitates comparisons of patterns (Figure 1.5). These figures show that in the four countries, the cut-off year seems to be associated with a quicker drop in fertility between the year just before the cut-off year and the first year of the health module. This is also visible in Colombia, but the drop is slight.

Figure 1.5: patterns of relative TFRs around the cut-off year of the health module in four DHS



Such fertility trends were computed for 191 surveys. The average for the 191 countries is shown on The degree of distortion varies greatly across surveys. Cluster analysis was used to group surveys according to the pattern of fertility in the years preceding and following the cut-off year. Four broad groups of surveys can be distinguished.

- The first group represents very distorted patterns, and includes 27 surveys, most of them (23) from sub-Saharan Africa, three from Asia (including the 1990 Pakistan survey), and one from the MENA region (Yemen). On average, the relative fertility level drops from above 1.2 the year before the health module to around 0.8 the cut-off year.
- The second group also shows distorted patterns, but the impact of the health module is less pronounced than in the first group. It mainly includes surveys from sub-Saharan Africa (29 surveys out of 34), as well as a few Latin American surveys (Haïti, Guatemala) and surveys from MENA and Asia.
- The third group also mainly includes sub-Saharan African surveys (15 out of 24). As the second group, it also shows distorted patterns, and is characterized by an overall decrease in fertility. It

also includes surveys from Latin America (Brazil, Bolivia, Nicaragua), Asia (Bangladesh, India) and MENA (Jordan)

- The fourth group is the largest, and comprises more than half of the surveys. On average, the impact of the health module is visible but moderate. This group includes the three European surveys, most of the Latin American and Asian surveys, as well as more than half of the MENA surveys. In contrast, only a quarter of the sub-Saharan African surveys are in that group.

Figure 1.6. This suggests that – on average – the health module is clearly associated with a disruption in the level fertility. The relative fertility level drops from 1.08 just before the cut-off year to 0.90 the first year of the health module, and remains at that level for three years.

The degree of distortion varies greatly across surveys. Cluster analysis was used to group surveys according to the pattern of fertility in the years preceding and following the cut-off year. Four broad groups of surveys can be distinguished.

- The first group represents very distorted patterns, and includes 27 surveys, most of them (23) from sub-Saharan Africa, three from Asia (including the 1990 Pakistan survey), and one from the MENA region (Yemen). On average, the relative fertility level drops from above 1.2 the year before the health module to around 0.8 the cut-off year.
- The second group also shows distorted patterns, but the impact of the health module is less pronounced than in the first group. It mainly includes surveys from sub-Saharan Africa (29 surveys out of 34), as well as a few Latin American surveys (Haïti, Guatemala) and surveys from MENA and Asia.
- The third group also mainly includes sub-Saharan African surveys (15 out of 24). As the second group, it also shows distorted patterns, and is characterized by an overall decrease in fertility. It also includes surveys from Latin America (Brazil, Bolivia, Nicaragua), Asia (Bangladesh, India) and MENA (Jordan)
- The fourth group is the largest, and comprises more than half of the surveys. On average, the impact of the health module is visible but moderate. This group includes the three European surveys, most of the Latin American and Asian surveys, as well as more than half of the MENA surveys. In contrast, only a quarter of the sub-Saharan African surveys are in that group.

Figure 1.6: Average pattern of relative TFRs around the cut-off year of the health module in 191 DHS

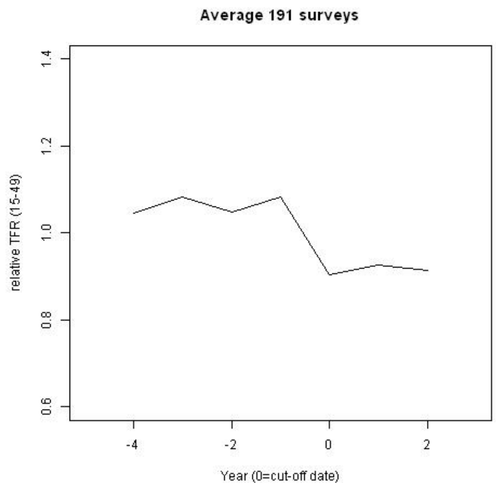
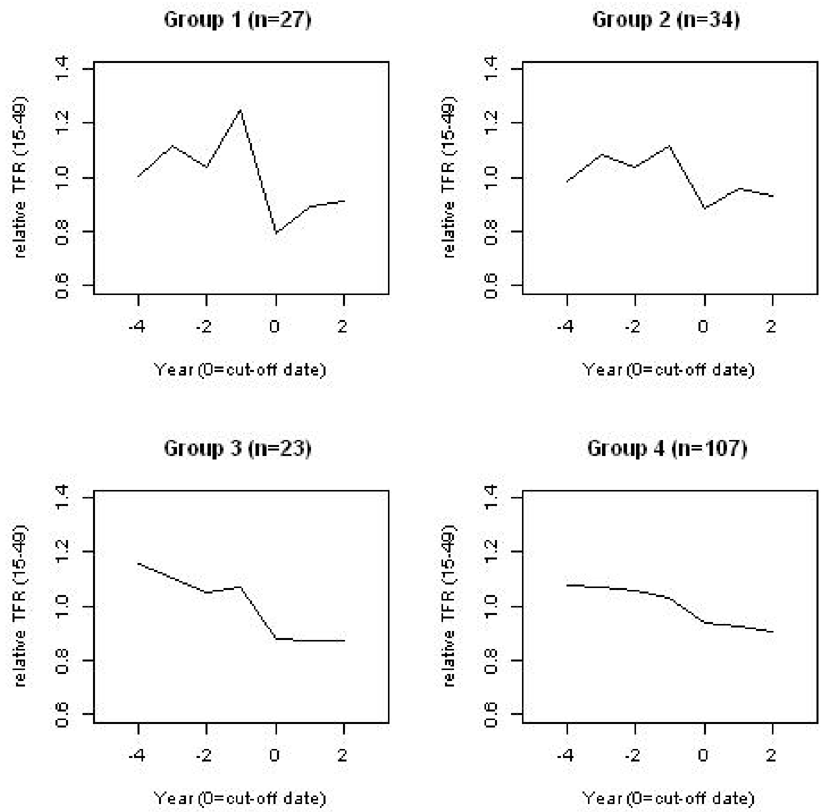


Table 1.3: Distribution of surveys by region and groups of patterns of fertility around the cut-off year.

Region	Group 1	Group 2	Group 3	Group 4	Total
Sub-Saharan Africa	23	29	15	33	100
Latin America	0	3	3	27	33
MENA	1	1	2	13	17
Asia	3	1	3	31	38
Europe	0	0	0	3	3
	27	34	24	107	191

Figure 1.7: patterns of relative TFRs around the cut-off year of the health module in four groups of surveys



1. The average pattern was also computed by regions (

Figure **1.8**). This confirms that sub-Saharan Africa is the region where the patterns are most distorted. The situation is better in Latin America and Asia, but no region is exempt from distortions. Comparisons of average patterns by DHS phase also indicate that this is an enduring characteristic of the DHS (Figure 1.9).

Not surprisingly, there is a correlation between the distortions of the TFR and the underestimation of fertility. The surveys affected by a large underestimation (

Table 1.1) are over represented in the group 1 and group 3, in which distortions are the most pronounced. Two-thirds of the surveys affected by large underestimation are in these two groups. These two groups include only 40% of the surveys with moderate underestimation. The average level of underestimation is higher in groups 1 and 3, and lower in groups 2 and 4.

Figure 1.8: patterns of relative TFRs around the cut-off year of the health module in four regions

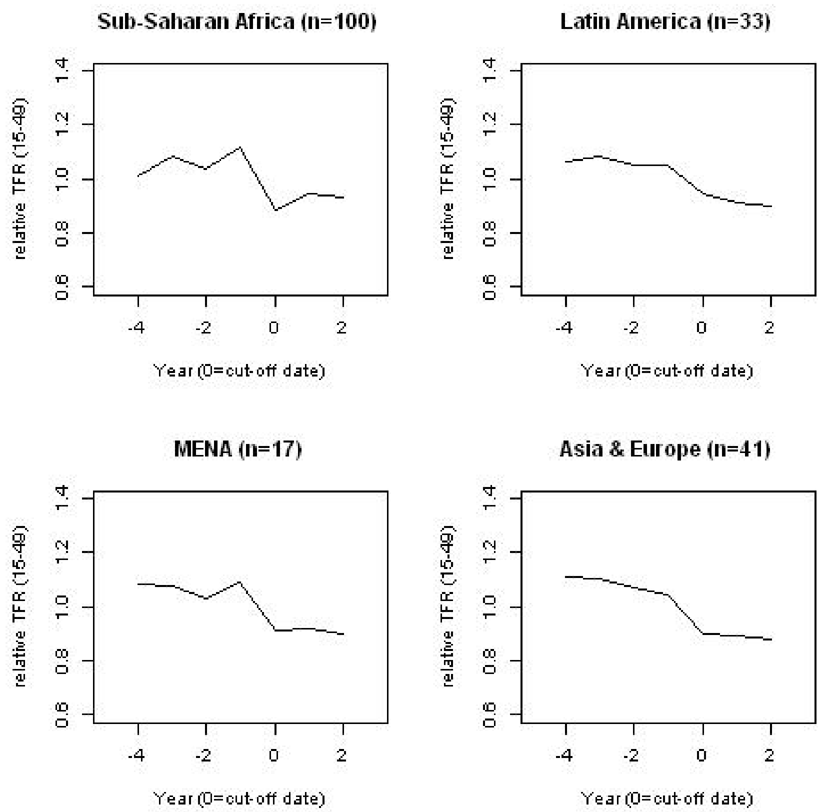
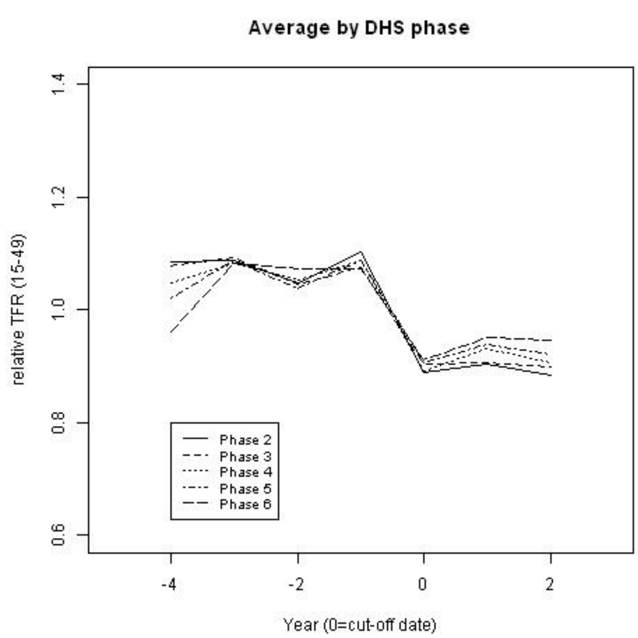


Figure 1.9: Average pattern of relative TFRs around the cut-off year of the health module in 191 DHS, by phase of survey



Part 2. Evidence that increased survey complexity has degraded the quality of the birth histories[9]

2.1. Introduction

During the past quarter-century, governments and funding agencies have increased the scope of the DHS, which now includes questions about nutrition, disease prevalence, sexual risk behaviors, knowledge about HIV/AIDS, and in some countries, detailed questions about domestic violence, maternal mortality, and biomarkers including blood collection for anemia and HIV testing.

This expansion has been hypothesized to affect data quality: the increase in complexity and length of the DHS questionnaire is likely to result in decreased training time and attention devoted to the original elements of the core questionnaire, specifically the birth history, which is used as the basis for both fertility and child mortality estimates. Such reduced attention is especially likely if, as suggested by anecdotal evidence from DHS staff, the duration of fieldworker training has not increased at the same rate that the questionnaire has grown, resulting in less time spent per section of the questionnaire.

The increase in the length and complexity of the DHS questionnaire comes from two sources: one, questions added to the core DHS questionnaire (which is revised with each five-year survey phase) such as detailed information on child feeding, which are asked in every country. Two, surveys can incorporate additional modules that are not part of the standard DHS core questionnaire, generally added at the request of the host country government. Common modules include the domestic violence module, maternal mortality module, anemia biomarker testing and, in countries with high suspected HIV prevalence, HIV biomarker testing. When such modules are added, fieldworker training must cover these detailed questions, including special protection procedures for respondents who answer the domestic violence questions, special informed consent procedures for blood collection to test HIV seroprevalence and, if the interviewers themselves are collecting the blood, training on biomarker collection. In surveys with longer core questionnaires plus additional modules, the burden on the interviewers is even higher, and any problems with data quality are potentially exacerbated.

In this analysis, I address the question of whether increasing the length and complexity of the DHS survey instrument leads to poorer data quality and thus biased mortality estimates. I first explain the likely causes and consequences of one measure of data quality: displacement of births out of the five years prior to each survey. I then examine differences in displacement by DHS survey characteristics, including core questionnaire length and implementation of modules including HIV biomarker testing. To document these patterns, I analyze every available DHS since the inception of the project, for a total of 198 surveys,[10] and use the survey as the unit of analysis.

[9] Sarah E.K. Bradley is the principal author of this section of the paper.
[10] Surveys that did not include a boundary year for the maternal and child health sections of the questionnaire are excluded from analysis.

2.2. Data and methods

Why expect displacement?

In the DHS questionnaire, the first section after the respondent's demographic characteristics is a complete birth history, recording the dates of birth and, if the child has died, date of death, for every child a woman has ever delivered. In later sections of the questionnaire, women who have given birth in the past five calendar years are asked a long series of questions about the prenatal care, delivery, postnatal care, vaccination record, recent illnesses, and feeding practices for their youngest child. Many of these questions are repeated for each child born in the five calendar years prior to the survey, so that a woman who was interviewed in 2006 and had given birth three times since January 1, 2001[11] would be asked many of the same questions three separate times. Savvy interviewers—and, perhaps, interviewed women[12]—quickly learn that the questionnaire can be shortened by "reducing" the number of births that occurred in the past five years by displacing births over this "boundary year." With more and more questions added to the child health sections in recent years, the incentive to displace births has increased.

<Figures 1 and 2 about here>

In countries with stable or declining fertility rates, we would expect to see the number of births recorded in each calendar year to be roughly the same, or slightly lower, as time progresses. Figure 2.1, however, shows a spike of births in 2000, followed by a sharp drop in the number of births in 2001, and then resumption of a relatively flat trend in 2002 through 2005.[13] Though some of this heaping on 2000 likely reflects digit preference for round numbers, it is unlikely that such displacement can be explained by heaping alone. A much more plausible explanation is that interviewers are selectively displacing births out of the five-year reference period, across the "boundary year" of 2001 and into 2000.

Consequences of displacement

This displacement affects the estimation of both fertility and mortality rates. Both rates use data from the most recent time period: the five- (or, depending on the level of estimation and degree of precision required, a three- or ten-) year period prior to survey. Births in this time period are the numerator for age-specific and total fertility rates, and if births are displaced out of the recent time period and into the prior one, the numerator will be reduced and fertility rates underestimated.

[11] In some surveys these questions applied to the most recent 3 years; this is accounted for in analysis.
[12] Because birth dates are recorded in an earlier section of the questionnaire, however, it is unlikely that interviewed women would be able to recognize this pattern and change the reporting of their children's birth dates, unless there are multiple women interviewed in the household and women have observed each other's interviews. Because most households only have one eligible woman (ages 15-49), any interviewee-initiated displacement is likely minimal.
[13] Because data for 2006 are incomplete, this graph is truncated at 2005.

In infant and child mortality rate calculation, the number of children surviving until the beginning of the age interval during the period of observation form the denominator, and the number of children who died in the age interval form the numerator (details of this calculation are described in Rutstein and Rojas 2003). If displacement affected surviving and deceased children equally, mortality estimates would be largely unaffected by displacement. As Figures 2.1 and 2.2 (Figure 2.2 is a replicate of Figure 2.1 on the log scale to highlight differences) show however, the birth dates of deceased children are displaced at a higher rate than those of living children. This differential displacement is understandable for two reasons: one, dates of birth for deceased children are more difficult to recall than for living children because the date of birth cannot be approximated from the child's current age. Two, interviewers seem reluctant to ask women questions about their dead children's prenatal care, delivery, birth weight, and duration of breastfeeding. Despite the fact that there are fewer questions asked about dead children than surviving children (feeding practices, child health status, and vaccination history are skipped for dead children), interviewers are still more likely to displace the birth dates for deceased children than for surviving children, as demonstrated below. In addition to levels of displacement, patterns of displacement also appear to differ by child survival: as shown in Figure 2.2, the sharp decline in the number of births in the boundary year for surviving children is not found for deceased children. Instead, it appears that for deceased children, the entire level of reported births is lower in the five years before the survey than in the prior five year period, with a spike in 2000. It therefore appears that for deceased children, births are displaced not only from the boundary year, but from the entire five year period prior to survey, and moved into the boundary year -1. It is also possible that such births have been omitted entirely, which would be much more difficult to detect. Displacement, however, is clearly detectable and measurable, as described below.

This differential displacement has clear consequences for mortality estimation, especially because a standard way to examine trends in mortality is to calculate and compare under-five mortality rates in the 0-4 years, 5-9 years, and 10-14 years prior to survey, using birth histories collected in a single survey. Though there are clear limitations to this approach due to truncation of maternal age[14] and increased exposure to the risk of mortality for births further back in time, this approach is widely used to proxy mortality trends. The displacement shown in the figures above will clearly underestimate mortality in the most recent period and overestimate mortality in the prior period. Comparing these estimates to each other would, in turn, overestimate the downward slope of any real decline in child mortality.

Measuring birth displacement

To identify the level of displacement in each survey and by survey characteristics, I create a **boundary ratio** defined as

$$100 * (B_{boundary\ year}/B_{boundary\ year\text{-}1})$$

[14] DHS surveys only interview women 15-49. In periods 10-14 years in the past, interviewed women were 1-29 years old. Thus any births that occurred when mothers were ages 30+ will be omitted from estimates for earlier time periods. Recall of the timing of births and deaths may also be less accurate for periods further back in time.

Where B is the reported number of births in the calendar year, and the boundary year is defined in the survey questionnaire, usually as the year of survey minus five.[15] If births were evenly distributed across years, the measure would equal 100. Minor fluctuations are expected due to random noise, but unless birth rates are changing quite dramatically, the number of births in two adjacent years should be approximately equal. Ratios under 100 therefore indicate displacement or bias. I calculate ratios for all births, and separately for surviving and deceased children. The difference in these ratios indicates the amount of differential displacement of deceased children compared to surviving children.

Survey characteristics

To see if, as hypothesized, data quality (as measured by birth displacement) changes with the survey instrument, I use several measures of survey length and complexity. As a proxy for the number of questions in the DHS questionnaire (only obtainable by hand counting), I use the DHS survey phase. The approximate number of questions asked to the average woman (a sexually active married woman with one birth in the last five years) changed across survey phases (Macro international 2009):

- Phase 1, 1984-1989: 205 questions
- Phase 2: 1989-93: 266 questions
- Phase 3: 1993-97: 258 questions
- Phase 4: 1997-2003: 292 questions
- Phase 5: 2003-08: 439 questions
- Phase 6: 2008-13: 358 questions[16]

These are only rough approximations. Countries can, and do, add their own questions to the core questionnaire, and the actual number of questions asked depends in large part on women's responses and resulting skip patterns (i.e., women who have not heard of HIV are skipped out of all questions about HIV). To account for this variation, I also test an alternative measure: the median duration of interview, created from subtracting the start and end times of individual interviews.[17] This measure is used in tandem with two other indicators: the percentage of interviews that required a return visit (because durations cannot be calculated for interviews that were conducted over two sessions, as multiple start and end times cannot be recorded), and the average number of births per woman that occurred in the past five years, because of the repeated questions asked for each birth in this period.

[15] Curtis used a measure with the same numerator, but used as the denominator the average of $B_{boundary\ year-1}$ and $B_{boundary\ year+1}$ (Curtis 1995). This averaged denominator would be appropriate if we anticipated the births were heaped on the boundary year, rather than displaced only from one direction. As I see no evidence of displacement from the earlier time period into the boundary year, this averaged measure appears to conceal some of the displacement, and I believe the boundary ratio used here to be superior.

[16] Number of questions for Phases 1-5 are from Macro International 2009; Phase 6 estimate from author's count.

[17] I remove from this calculation all durations under 10 minutes, including negative values, as implausible and likely due to coding errors. In four early surveys, the start and end times were not recorded and durations not be calculated.

To see if the addition of modules affects displacement, I create dummy variables for whether the survey included the four most common (and potentially burdensome) modules: HIV biomarker collection, anemia biomarker collection, domestic violence, and maternal mortality modules. I also disaggregate data by region of survey, which accounts for a great deal of variation in date reporting. In the multivariate regression models shown below, I include several controls relating to the accuracy of date reporting. It is widely believed (at least within the DHS) that interviewers are more likely to displace births of less educated women, in large part because such women are less likely to report the exact dates of birth of their children (Pullum 2006), giving the interviewer a larger role in determining the child's birth date. To adjust for this possibility, I include as a covariate the percentage of women with no education in each survey. Finally, if a child's birth date is not recorded, the data are imputed using a standardized hotdeck-style procedure described in Croft 1991. To account for any differences brought about by such imputation, I also control for the percentage of births for which any part of the birth date (day, month, or year) are imputed.

Summaries of these characteristics are shown in Appendix Tables 1 and 2. In brief, the median duration of interview varies as expected with survey phase: the average survey took half an hour to complete in the 1980s, and close to one hour in 2003-2008, when the core questionnaire was at its longest. The median survey duration is also higher with implementation of each module. In surveys that included HIV biomarker testing, for example, the median interview is 10 minutes longer than surveys that excluded this component (53 vs. 43 minutes). The percentage of interviews requiring a second visit is also higher with implementation of each module. The average number of births in the past five years indicates the level of fertility, ranging from 0.56 in Latin American and the Caribbean to 0.8 in sub-Saharan Africa.

2.3. Results

As shown in Table 2.1, there is great variation in data quality, as measured with the boundary ratios described above, by survey characteristics. Boundary ratios were closest to 100, indicating the least amount of bias, in the earliest DHS phase, which also had the shortest questionnaire. The total boundary ratios (column 2) decrease fairly linearly with increasing survey phase until the most recent phase, in which the core questionnaire was reduced in size. It is important to note, however, that data from this survey phase are still being collected, and thus we are not assessing all data from Phase 6.

<Table 2.1 about here>

Consistent with other analyses of data quality (Curtis 1995; Pullum 2006), the boundary ratios detect the least amount of bias in the Latin American and Caribbean surveys, and the most bias in sub-Saharan African surveys. Data quality appears lower in surveys that included HIV testing, but this could simply be because most of the surveys that included HIV testing were in sub-Saharan Africa. Boundary ratios are lower in surveys that included anemia testing and the maternal mortality module than in those that did not, but surveys that included the domestic violence module appear less biased, according to the boundary ratio measure, than surveys that did not ask these questions. Again, this is probably related to the regional distribution: nearly 30 percent of surveys with the domestic violence module were conducted in Latin America and the Caribbean, compared to 14 percent of surveys that did not include the module. The over-representation of higher-quality Latin America and the Caribbean surveys in the domestic violence group likely explains this anomaly.

Columns 3 and 4 disaggregate the boundary ratios by the survival status of the child, and column 5 shows the difference in these two ratios. This difference is rather damning of the birth history data quality: on average, there is a gap of 20 percentage points between the ratios for surviving and deceased children, indicating not only a fair amount of displacement but also strong differential displacement by child survival. In some cases, this difference is likely the result of small sample sizes: in the Armenia 2010 survey, for example, this gap is 67 percentage points, but this is only because of the small total sample and low child mortality resulting in an absurdly low boundary ratio of 22 for deceased children. In many cases, however, this marker of poor data quality cannot be explained away as a statistical quirk. The Mozambique 1997 survey had a sample of almost 30,000 births, and an estimated under-five mortality rate of 201 deaths per 1,000 live births. The total boundary ratio for this survey is 66, and among deceased children, the boundary decreased to 36. These clear markers of poor data quality must result in poor estimates of mortality and fertility.

<Table 2.2 about here>

In Table 2.2, I examine factors associated with the total boundary ratio (all children). Regressions on the boundary ratios for deceased and surviving children (not shown) show largely the same results. Without controlling for other factors, the inclusion of HIV biomarkers in a survey does increase the degree of bias, reducing the boundary ratio from 88 to 84 as shown in Table 2.1. Similar results are seen for the inclusion of the maternal mortality module, but no statistically significant effects are seen with the domestic violence or anemia biomarker modules. Compared to Phase 1, all subsequent phases show significantly higher levels of bias, particularly Phases 4 and 5, which decrease the boundary ratio by 10 points compared to Phase 1. As expected based on Table 2.1, data quality is significantly worse in sub-Saharan Africa than in Latin America and the Caribbean. In univariate analysis, the average number of births in the five years prior to survey, percentage of interviews requiring revisit, and percentage of births with imputed information are all inversely associated with data quality, but none of these relationships remain statistically significant after controlling for other factors. Surprisingly, the median duration of interview is not statistically significantly related to the boundary ratio in any model, though this could be due to the aforementioned measurement error.

In multivariate models, the only statistically significant predictors of bias are the survey phase, location in sub-Saharan Africa vs. Latin America, and percentage of women with no education. HIV biomarker testing, which I hypothesized would be the strongest driver of poor data quality due to the increased burden, does not remain significant after controlling for region. Subsequent analyses (not shown) found no detectable difference in levels of bias by inclusion of HIV testing, even within region or survey phase, there was though it is possible that such a difference exists and I am underpowered to detect it. More plausibly, it seems that decreases in data quality are due to multiple factors, and that inclusion of HIV is simply one additional component that contributes to the overall survey burden.

2.4. Conclusions

Despite the DHS's well-earned reputation for excellent data quality, this analysis highlights serious problems with the birth history data. Moreover, these results demonstrate that such problems have been increasing as the survey instrument has increased in length and complexity, and that bias is greatest in sub-Saharan African surveys, net of other factors. This is especially troubling given the great amount of attention paid to recent declines in child mortality especially in this region. As described above, the bias detected in this analysis is quite likely to bias estimates of both fertility and mortality downwards. If further analyses demonstrate that apparent declines have been overstated due to data quality problems, the future of the DHS project, and thus the key source of demographic data for multiple countries, could potentially be in jeopardy.

Fortunately, the fact that the causes of this displacement are (presumably) well known means that they can likely be remedied. Questionnaire changes, such as asking the child health questions for the two most recent births regardless of when the children were born, would eliminate the boundary year and thus the incentive to displace births.[18] In addition, or as an alternative, estimation strategies could be altered, to use a shorter reference period, as implemented by the UN IGME when sample sizes allow. Neither of these changes would completely eliminate the problems. Interviewers will still want to shorten their burden, and especially avoid asking mothers about their dead children, even if the clear incentive identified by the boundary year is removed. Additionally, if births are displaced by more than one year (as seems likely for deceased children in the Uganda example), shortening the reference period will not avoid bias entirely. But such changes, in addition to decreasing questionnaire length and increasing fieldworker training, would go a long way in improving the quality of vital estimates.

I began this analysis wondering if there was a tradeoff between the quantity and quality of data collected. Though not causal, this analysis presents abundant evidence that increases in survey length and complexity are indeed strongly associated with reduced data quality. DHS has already taken steps to reduce the length of the core questionnaire in Phase 6, which seems to result in less bias. This analysis provides evidence that these changes are indeed moving in the right direction, and indicates that further reductions in the quantity of questions asked will continue to improve the quality of DHS data in the future.

Note: The figures and tables for this section are omitted. Pdf versions can be obtained from Thomas Pullum or Sarah Bradley.

[18] Thanks to Ron Lee for this suggestion.

References

Croft, T. 1991. Date Editing and Imputation. Demographic and Health Surveys World ConferenceProceedings, II: 1337–1356, Columbia, MD: IRD/ORC Macro.

Curtis, S.L. 1995. Assessment of the Quality of the Data Used for the Direct Estimation of Infant and Child Mortality in the DHS-II Surveys. Occasional paper 3. Calverton, MD: Macro International Inc.

Macro International. 2009. DHS questionnaire revision process. Presentation to USAID. Washington, DC.

Pullum, T.W. 2006. An Assessment of Age and Date Reporting in the DHS Surveys, 1985-2003.Methodological Reports No. 5. Calverton, MD: Macro International Inc.

Rutstein, S.O. and G. Rojas. 2003. Guide to DHS statistics. Calverton, MD: ORC Macro.

United Nations Inter-agency Group for Child Mortality Estimation (UN IGME). 2012. Levels and trends in child mortality. New York, NY: UNICEF.

Part 3. Sensitivity of fertility and mortality rates to omission and displacement in the birth histories[19]

3.1 Introduction and strategy

The birth history collected from each woman age 15-49 is essentially a listing of all the live births that the woman has had. The most important attributes of each birth, apart from the fact of the birth, are the month and year of the birth; whether the child has survived to the date of interview; and, if the child has died, the age at death. The month and/or year of the birth are sometimes missing, but if they are missing (or if, say, two births are impossibly close together), DHS will impute them. The imputation procedure is not believed to induce any kind of bias and will not be discussed here. The woman's own month and year of birth (sometimes imputed) and the month and year of the interview are used to calculate the woman's age at the time of the birth and the elapsed time before the survey, respectively. Age-period-specific fertility rates are calculated on the basis of the woman's exposure—the length of time when she was in an interval of age within an interval of time--and a count of the births she had within that window of exposure.

Age-period-specific mortality rates of children are calculated from a file of births, rather than women. Exposure (or risk) is expressed in terms of the length of time in an interval of age within and interval of time, and the outcome is binary—whether the child died or survived—rather than a count. The calculation of a rate is complicated by the fact that if the child died, we do not know the month and year of the death; we only know the age at death. This fact often leads to some ambiguity, which is resolved by allocating the risk and outcome to each of the two time intervals that the age interval may straddle. The standard under-five death "rates" are actually probabilities (or estimated probabilities) rather than rates, but we shall follow the convention of referring to them as rates.

If the birth histories include errors, then the rates *may* be incorrect, but that is not always the case. For example, displacement of birthdates within an interval of age and time will have no effect on fertility rates. Omission of children who died will have a far greater impact on mortality rates than on fertility rates; in an age interval with low mortality, the impact on fertility rates may be negligible.

This section of the report attempts to articulate the birth histories with the calculated rates with a sensitivity analysis. The strategy can be summarized as follows.

Step 1. Calculate a standard set of fertility and child mortality rates using the original data.

Step 2. Compare the events in the birth history with a set of criteria that are believed to characterize "correct" data.

Step 3. Adjust the events in the birth history to bring them into compliance with the criteria in step 2. This adjustment is accomplished by re-weighting the births.

Step 4. Re-calculate the standard set of rates using the re-weighted data.

[19] Thomas Pullum and Stan Becker are the principal authors of this section of the report.

This strategy produces two different sets of indices of data quality. The first set, coming from steps 2 and 3, measures the differences between the observed birth histories and the criteria. This set of indices can be described as levels of omission and displacement, i.e. errors in the birth histories. The second set, which can be described as errors in the rates, comes from steps 1 and 4, measuring the difference between the rates before adjustment and after adjustment. The errors in the rates will be compared with the statistical standard errors of the rates, adjusted for the survey design.

It must be understood that the terminology used here implies some assumptions that may or may not be valid. Certainly, for example, the term "errors" should be qualified to be "potential errors". "Criteria" are "plausible criteria". We can never know what the true or correct rates are, and in any case the DHS rates are calculated from survey data and are subject to sampling error as well as to a range of non-sampling errors, some of which are virtually impossible to identify.

3.1 Criteria for relationships within the birth histories

In the implementation of step 2 of the strategy, six criteria will be employed. Each criterion can be stated as a hypothesis. Other criteria could probably be added to the list. Three criteria relate to omission of births (particularly births that resulted in child deaths) and three relate to displacement of dates of births.

Omission Type 1. Sex-specific omission of births. It is hypothesized that that when the sex ratio at birth deviates from an expected value, it is because of a tendency to omit boys or to omit girls. The hypothesis is stated in a manner that is neutral with respect to whether the omitted births tend to be boys or girls. If the observed proportion of births who are girls is lower than expected, girls will be added with the weight adjustment procedure. Alternatively, if the proportion who are boys is lower than expected, boys will be added. Based on high quality international data, it is expected that the sex ratio at birth is 104 boys to 100 girls. In a variant, the criterion for the sex ratio at birth is set at 103 in Sub-Saharan Africa and 105 elsewhere.

Omission Type 2. Sex-specific omission of neonatal deaths. This hypothesis is also neutral with respect to whether the omitted deaths tend to be boys or girls. Either boys or girls will be added in order to bring the balance to the hypothesized level. The additions will increase the number of births as well as deaths. It is expected that the sex ratio of neonatal deaths is 150 boys to 100 girls, i.e., that the proportion who are girls is 0.40 and the proportion who are boys is 0.60.

Omission Type 3. Omission of neonatal deaths. It is hypothesized that births which resulted in neonatal deaths tend to be omitted, leading to an under-estimate of fertility and of neonatal, infant, and under-five mortality. Very few questions in the core DHS questionnaire are asked about children who had died at any age before the survey. Nevertheless there is probably some reluctance of respondents to mention children who died, especially if they were very young when they died, as well as a some reluctance of interviewers to probe for such births. The criterion for detecting such omissions is not a fixed number, but is more complex, and will be described separately.

We also hypothesize three types of potential displacement. Displacement is much different from omission because it can be more readily observed as an irregularity in a distribution. The expected distribution is simply a statistical smoothing of the observed distribution.

Displacement Type 1. Heaping of age at death at 12 months. In all surveys, children's deaths are disproportionately reported at age one year or 12 months. The period of infancy is the first year following the birth, i.e. completed months 0-11. The response "12 months" is interpreted as 12 completed months, and is therefore more than one year after the birth. Deaths are re-distributed in a range that includes 12 months but extends to two months before 12 and two months after 12. The effect of the re-distribution is to increase the infant mortality rate, to decrease the rate for ages 1-4, and to leave the rate for 0-4 unchanged.

Displacement Type 2. Heaping of age at death at 7 days. A similar type of heaping is observed at 7 days. The early neonatal mortality interval is the first seven days after birth. Early neonatal deaths are defined to occur during the first week. Seven completed days would place the death in the post-neonatal interval. Deaths are re-distributed in a range that includes 7 days but some days before and after 12. The effect of the re-distribution is to increase the early neonatal mortality rate, to decrease the late neonatal rate, and the leave the neonatal rate and all other rates unchanged. The standard list of under-five rates produced by DHS does not include the distinction between the early and late neonatal intervals. This displacement is included in anticipation of increased emphasis on that distinction in the future.

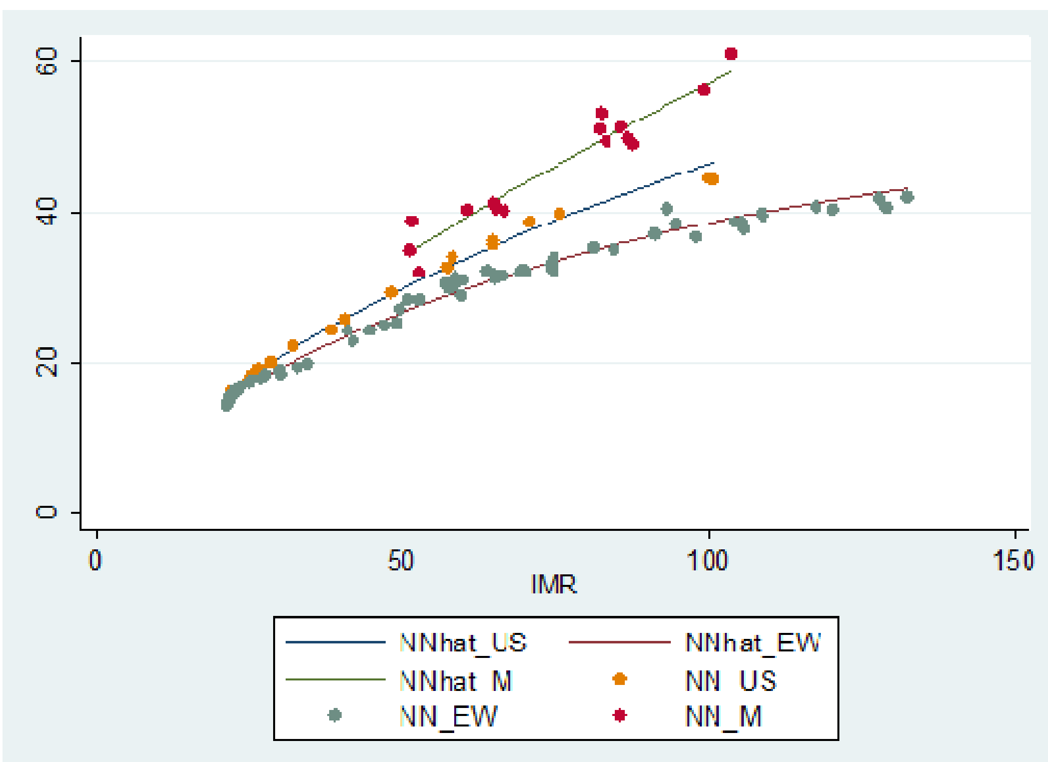
Displacement Type 3. Types 1 and 2 refer to heaping, in which cases tend to be moved toward a number that represents a rounded response. Type 3 is much different. It refers to a tendency to move events in one direction rather than symmetrically. It is hypothesized that births which actually occurred after the beginning date for the child health questions will tend to be moved to an earlier date, on the other side of the threshold. This kind of displacement will substantially reduce the number of questions that the interviewer must ask about the child. The start date is generally (but not always) January of the fifth calendar year prior to the first month of fieldwork. That is the date when the extra questions begin to apply. Later, during data processing, some data are discarded so that the reference period for the child questions is always the five years (60 months) prior to the month of interview, with a starting month that is not fixed but varies from child to child, depending on the month of interview.

Corrections of the data are done with an iterative procedure that cycles through the six adjustments in succession, until convergence is reached and all corrections are achieved in a mutually compatible manner. It is our experience that the procedure always converges, within about six iterations, to any convergence criterion. The sequence of the adjustments does not matter. The iterative procedure forces the data to match all the criteria, regardless of how close the original data may have been. That is, even if the original data were very close to the criteria, the criterion will be applied—and will make almost no difference.

3.2. Further detail on omission of neonatal deaths

The criterion for omission type 3 is derived from a relationship identified by Hill and Choi (2006). Using historical data from England and Wales, they found that when the IMR is greater than 20 deaths per 1000 births, the ratio NN/IMR fits almost perfectly a linear regression on log(IMR). (Here IMR is the infant mortality rate and NN is the neonatal mortality rate. Hill and Choi did not attempt to motivate this relationship with a model, nor will we, but it is an empirical regularity that appears in other data, including the United States in the early 20th century and Matlab. Figure 3.1 shows the observed and fitted values of NN and IMR from these three data sets.

Figure 3.1. Observed and fitted values of the neonatal mortality rate (NN, the vertical axis) and the infant mortality rate (IMR, the horizontal axis), derived from the linear regression of NN/IMR on log(IMR), for IMR>20, using data from England and Wales (E&W), the United States (US), and Matlab (M).



In the original regression of Hill and Choi, the intercept is 1.37 and the coefficient of log(IMR) is -0.214; R2 is 0.98. (These numbers were not in the Hill and Choi paper but were calculated from data kindly provided by Ken Hill.) In the U.S. data, the intercept is 1.34 and the coefficient of log(IMR) is -0.190; R2 is again 0.98. In the Matlab data, the intercept is 1.30 and the coefficient of log(IMR) is -0.158; R2 is a less impressige 0.55. In our analysis, numbers in the ranges of these three settings are being applied. The relationship is being re-written by expressing IMR as NN+PN, where PN is the postneonatal mortality rate, leading to NN as an implicit function of PN. If we make the assumption that the postneonatal mortality rate is accurate, we obtain leverage through this relationship to estimate what the neonatal mortality rate should be. In order to achieve that level (if the observed value is too low), we add children who will then appear both as births and as neonatal deaths.

Because NN is an implicit function of PN, rather than an explicit function, it is necessary to apply an iterative Newton-Raphson procedure to get a fitted value of NN, given the observed value of PN.

Hill and Choi (2006) also identified an empirical relationship that could allow separate estimation of the early neonatal mortality rate using PN, but we are not using that relationship in this analysis.

3.3. Further detail on redistributing displaced births

Identifying and correcting for all three types of displacement is done in a very similar way. The data are collapsed (i.e. the weighted frequencies are summed) within a specific interval of five intervals of age or four intervals of time. A poisson regression is fitted to the observed counts, with age or time as a linear covariate. The fitted values will have the same total as the observed values, and the logs of the fitted frequencies will be linear. This is a simple method for maintaining a total count but having a smooth progression from one interval to the next.

3.4. Correcting for omissions and transfers by altering the weights

To assess the impact that the six types of errors have on the fertility and mortality rates, we will artificially correct the data, and compare the rates calculated for the corrected data with the original rates, which were calculated without the corrections. This section will describe the procedure for making these corrections.

If there is evidence that some types of births--for example, births that resulted in a neonatal death--tended to be omitted, a possible strategy to correct the data would be to add some artificial cases to the data file to compensate for the omissions. If there is evidence that some type of birth--for example, a birth that occurred soon after the beginning of the time interval for the health questions--tended to be displaced, a possible strategy would be to artificially change the birthdates for some of the births in the interval with an excess and move them into the interval with a deficit.

Unfortunately, it is difficult to add cases or to alter the dates for specific cases. Adding artificial cases with a neonatal death, for example, would require imputing the b variables other than the ones for age at death. Shifting birthdates would require some random or arbitrary procedure to select the specific cases to shift, making it very difficult for anyone to replicate the results. An alternative strategy will be used here, namely to manipulate the weight variable, v005, in a manner that simulates the addition or transfer of cases.

It may be helpful to summarize the construction and primary use of the DHS weight variable. DHS surveys have a complex sampling design that typically includes specifying strata (usually combinations of sub-national region and urban/rural residence), selecting clusters, randomly selecting households within clusters for a household survey, and then selecting all women in the age 15-49 for separate interviews that include birth histories (often there is also a survey men). The weight variable v005 on the individual woman's record is calculated to be inversely proportional to the net probability that a case in the population will be selected for the sample. The weight for each birth in the BR file is taken from the mother's weight in the IR file.

The main reason for using weights is to produce unbiased estimates of population characteristics. If the weights are ignored, then all estimates—proportions, means, regression coefficients, etc., will be biased toward the categories of women that are over-sampled, and away from the categories that are under-sampled. It is standard DHS practice to use weights for virtually all data analyses.[i]

The weight variable v005 is normalized to have a mean of one for each woman or birth.[ii] Under the pweight option, Stata always re-normalizes the weights to have a mean of 1. This default cannot be over-ridden, with potential implications that are taken into account here.

Our strategy will be as follows. First, if there is evidence of omission for a type of response, then the weight for all cases with that response will be increased by a multiplier to reach the target number of cases. Second, if there is evidence of displacement from one type or response to another, then the weight for all cases with the over-represented response will be decreased by a multiplier, and the weight for all cases with the under-represented response will be increased by a multiplier, calculated to achieve the target number of cases in the two categories. We will describe the process in more detail.

Define a generic variable, X, which can have non-missing values X=1, 2, or 3 and missing value *X*=".". For example, *X* could be age at death. The missing values would apply to surviving children, with *X*=1 for neonatal deaths, *X*=2 for post-neonatal deaths, *X*=3 for later deaths. The proportion of deaths with *X*=1, out of those with *X*=1 or 2, i.e. the proportion of infant deaths that are neonatal, is p. We have reason to believe that neonatal deaths tend to be omitted entirely, both as births and as deaths, and the proportion "should be" P. We want to increase the weights for those children with code *X*=1.

Say that *W*i is the total weight for those cases with *X*=*i*, with *W*4 as the total weight for the cases missing on *X*. Then we observe $p = W_1/(W_1 + W_2)$. We need a factor *f*, such that when the weights for cases with X=1 are multiplied by f, we will have $P = fW_1/(fW_1 + W_2)$. With some algebra it is found that $= \left(\frac{P}{1-P}\right) / \left(\frac{p}{1-p}\right)$. If *P*>*p*, which would be consistent with a hypothesis of omission, then *f* will be greater than 1. Thus, the individual weights for all cases with *X*=1 should be multiplied by *f*. No change should be made to the weights for cases with other values of *X*.

The increase in the total weight for the cases with *X*=1 will increase the total sample size. The normalized weights for the entire sample will be changed (that is, will differ from what they were before the weights for cases with *X*=1 were increased). The normalization itself will not be an issue for the calculation of any statistics, except that the additional cases are not "real". That is, the apparent increase in the normalized sample size, of $W_1(f-1)$ cases, is artificial. The effect should be negligible, but standard errors calculated from the artificially inflated sample should be adjusted upwards.

Next suppose that there is evidence that some cases have been transferred from one category to another. For example, it may be reasonable to hypothesize that $p = W_1/(W_1 + W_2)$ is too low, not because of omissions from category 1 but because deaths that should have been reported with *X*=1 were misreported with *X*=2. We need a strategy to increase the weight with *X*=1 and simultaneously reduce the weight with *X*=2 without altering the total weight for the two categories. In this situation, we calculate the expected values of the total weights W_1 and W_2 under some model and refer to them as E_1 and E_2, respectively. The sum of the *E*'s is equal to the sum of the *W*'s. Then in the individual-level file, each weight in group i (i=1,2) is multiplied by E_i/W_i, preserving the total and achieving the desired balance between the two groups. This approach can be expanded to any number of groups.

3.5 Calculation of rates using the adjusted data

The fertility and mortality rates relevant to this report are as follows:

Fertility: The standard set of seven age-specific rates for five-year intervals of age 15-19 through 45-49; and the Total Fertility Rate (TFR). These rates appear in the main report on every DHS survey.

Mortality: The standard set of five age-specific rates that appear in the main report on every DHS survey: neonatal; postneonatal; infant (the sum of neonatal and post-neonatal); child (for ages 1-4) and under-five (for ages 0-4). The infant, child, and under-five rates are actually estimates of the probabilities 1q0, 4q1, and 5q1, in conventional life table notation.

Time intervals: These rates are calculated for three intervals of time: 0-4, 5-9, and 10-14 completed years before the survey. DHS reports normally also give fertility rates for the three years before the survey, i.e. 0-2 completed years, but only for the national estimate. Fertility rates for subpopulations are only given for 0-4 years before the survey, and fertility trends are described with the rates for 0-4, 5-9, 10-14, and 15-19 years before the survey. DHS reports normally present the mortality rates listed above for 0-4 years before the survey. Mortality rates given for sub-populations are normally for 0-9 years before the survey. We will not go back more than 15 years before the survey because of the increasing selectivity of respondents who have survived and the increase in recall error.

The procedure described here is applied to the file of births, and it alters the weights attached to the births. The mortality rates for children are always calculated from this file, both before and after the adjustment to the weights.

It may not be clear how the fertility rates are calculated from the adjusted weights. It is the weights in the file of births, which is the source of the numerators of the fertility rates, that have been modified. However, nothing has been said about the weights in the file of women, which is the source of the denominators of the fertility rates, i.e. the exposure component of the rates. Those are left unchanged. It would appear to be inconsistent to have one set of weights (the original sample weights) for the denominators, and another set of weights (the adjusted weights) for the numerators. (We actually use poisson regression to calculate the fertility rates, and log probability regression to calculate the mortality rates, but the terminology of numerators and denominators is still applicable.)

Indeed, for the fertility rates, the re-weighting only applies to the numerators. Each birth, which would have a count of 1 with the original calculation, is replaced by the ratio (rather than a count) of the adjusted weight to the original weight. Then the original weights are applied to both the numerators and the denominators. In terms of the calculations, this will raise or lower the count of births according to the ratio of the adjusted weight to the original weight.

3.6 Conclusion

The results of this approach are not ready for inclusion in this paper but, as stated earlier, will appear shortly in a DHS Methodological Report.

Additional references not given earlier:

Guillot, Michel, Patrick Gerland, François Pelletier, and Ameed Saabneh. 2012. Child Mortality Estimation: A Global Overview of Infant and Child Mortality Age Patterns in Light of New Empirical Data. *PLOS Medicine*
http://www.ploscollections.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001299.

Hill, Kenneth, and Yoonjoung Choi. 2006. Neonatal mortality in the developing world. *Demographic Research* 14: 429-452

[i] For some purposes, data quality checks do not use weights. Also, as described in the Guide to DHS Statistics, some multivariate analyses conducted by DHS do not use weights, but this is now a minority practice within DHS.
[ii] The weights are then multiplied by 1,000,000 and rounded to the nearest integer, but the factor of 1,000,000 will be ignored in this discussion.