Quantifying and Correcting Bias in Fertility Estimation from Complete Birth Histories

Xie Rachel Kulikoff

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Public Health

University of Washington

2018

Committee:

Laura Dwyer-Lindgren

Nicholas Kassebaum

Program Authorized to Offer Degree:

Global Health

Abstract

Quantifying and Correcting Bias in Fertility Estimation from Complete Birth Histories

Xie Rachel Kulikoff

Chair of the Supervisory Committee:

Laura Dwyer-Lindgren, Assistant Professor

Global Health

Introduction

Accurate and unbiased estimates of fertility are critical in creating population policy and evaluating maternal mortality ratio. In settings without the routine collection of vital statistics, fertility data often comes from complete birth histories, which can suffer from biases due to the passage of time, including recall bias—women forgetting their births—and selection bias—women who die having systematically different fertility than those who live to be surveyed. Comprehensive estimates of time-related bias using nationally representative surveys in multiple locations do not exist.

Methods

This paper develops a method to measure and correct for these biases using 299 complete birth history surveys. I calculate cohort fertility rates of the same cohort from different surveys with different recall periods, find the ratio of a reference cohort fertility rate from the survey with the lowest recall to each other estimated cohort fertility rate, and model the log ratio as a function of recall length, reference recall length, age, maternal mortality ratio, female adult age-specific

mortality, and region. These ratios are then predicted as if the reference recall was 0 and applied to the raw cohort Age Specific Fertility Rate (ASFR) to get a value that is adjusted for bias.

Results

The cohort-recall specific estimates for each country of analysis showed considerable patterning across different recall periods. Predicted ratios of reference ASFR, with the lowest possible recall within a country, cohort, and age, to survey specific ASFR ranged between 0.66 for a three year recall of the 1948 birth cohort in Timor-Leste and 1.40 for a 26 year recall of the 1965 birth cohort also in Timor-Leste, with a mean and median of 0.98. The relationship between recall and the ratio varied by age. For younger ages, including 15-19 and 20-24 years, the model found an underestimation of fertility in relation to the reference with an increasing recall period; for older ages, including 25-29, 30-34, 35-39, and 40-44 years, the model found an overestimation of fertility in relation to the reference with an increasing recall period.

Discussion

This analysis represents the first systematic quantification and correction of fertility bias due to the passage of time. The results of this analysis suggest that selection bias due to differential fertility among those mothers who die and those who live to be surveyed contributes to the bias in fertility estimates. The application of the results could substantially increase the amount of data available to be used back in time if restrictions on recall are lifted. Limitations include the instability of the input data, the non-inclusion of age heaping and birth transference in analysis, and the focus on complete birth history methods to the exclusion of indirect methods. Additional work, including the development of alternate methods, is necessary to validate the results of this analysis.

Supplementary Material

Supplemental material includes graphs of cohort ASFR results for each country age survey.

Table of Contents

Introduction

Accurate fertility statistics are important in the evaluation of population policy and population estimation and projection. Fertility estimates are also an input to the calculation of maternal mortality ratio (MMR), thus, accuracy is paramount to evaluation of progress toward Sustainable Development Goal (SDG) 3.1, decreasing MMR to <70 per 100,000 live births by 2030.[1] In countries without routine registration of births through a vital registration system, the majority of fertility data comes from birth histories. Complete birth history (CBH) questionnaires, generally situated in the women's module of large survey series such as the Demographic and Health Survey (DHS), collect information from women of reproductive age (15 – 49), including their date of birth, the date of birth of each of their live born children, and health information on each child.[2]

CBH data can suffer substantially from various forms of bias, which can result in inaccurate fertility rates, birth counts, and MMR levels. These forms of bias, as summarized in the DHS Methodological Report 12[3], include:

1) misreporting of the date of birth of children,
2) omission of births,
3) displacement of births before a cutoff date (in order for a surveyor to avoid additional questions asked pertaining to children born in the 5 years previous to the survey),
4) selection bias due to differential fertility among women who die or internally migrate before the date of the survey,
5) misreporting of the age of the mother interviewed, and
6) the underestimation of the all-women correction, which is used in surveys when only ever-married women are surveyed.

Only a few of these sources of bias have been systematically evaluated. DHS conducted its own evaluation of bias due to misreporting of children's birth dates, omission of recent births (for the same reason that there is displacement of births before a cutoff date), and mistakes in the sampling scheme, by comparing reported fertility estimates to fertility estimates reconstructed using Poisson regression on age and period, finding that there was considerable heterogeneity in the agreement between the direct and reconstructed estimates, with some locations having very good agreement and others with poor agreement.[3] In addition to only focusing on a few sources of bias, this analysis also was limited by using only a smoothed regression model and assuming a constant age pattern of fertility.

To address the potential of biases increasing back in time, both DHS and the Institute for Health Metrics and Evaluation (IHME) restrict length of recall in the estimation of fertility from CBHs, with DHS restricting to 20 years and IHME restricting to 15 years. It will be useful to briefly describe the methods by which DHS and IHME make use of complete birth history data to come to estimates of age specific fertility rates.

Calculation of ASFR from Birth Histories

DHS and IHME both compute period Age-Specific Fertility Rate (ASFR) from raw CBH data. DHS obtains ASFR using births in the time period between 1 and 36 months previous to the date of the interview divided by the women-years of exposure in the same time period. Births occurring in the month of the interview are not included because they are censored by the date of the interview—that is, it doesn't represent the fertility behavior of the full month. DHS also offers guidelines for calculating trends in ASFR, which they define as ASFR for four- or five-year periods of time preceding the interview. The calculation of this period ASFR proceeds similarly to the calculation of current ASFR, except that it is done for 1-60 months, 61-120 months, and so on to 181-240 months before the survey, and the women-years of exposure within each period is 60 months for the 5-year period.[4] The information for each age group will be truncated as women age out of the group surveyed.

The methods IHME uses to process CBH data are similar in the definition of ASFR and the conceptual framework of number of births over person-time of exposure. The main difference between IHME methods and DHS methods is that IHME, in order to be consistent with child mortality processing, uses a 15-year recall period, and estimates are averaged over non-overlapping three-year periods and placed in the middle of the period to address stochasticity and small sample size. The exception is for older age groups at the time of the survey, for which IHME periodically decreases recall period so as not to introduce too much sampling bias. For example, for the age group 45-49, IHME does not use the maximum recall of 5 years, because it would only capture the fertility of women aged 49 at the time of the survey, as the other women aged out of the sample. The CBH data, processed in the way descried above, acts as one type of input data to IHME's fertility model.

By restricting recall period, both DHS and IHME are losing a substantial amount of potential data, especially for estimating ASFR for younger age groups, and neither make adjustments for fertility estimates up to their respective recall period restrictions. In this thesis, I aim to measure, quantify, and propose a correction mechanism for biases in fertility data due to the passage of time. This will allow for the incorporation of new data and ensure the quality of data currently being used. This kind of correction would not only potentially increase the quality of fertility data and estimates, but also could have far-reaching effects on the estimation of MMR, as well as IHME's analysis of the Socio-Demographic Index (SDI), which is a composite indicator including total fertility under 25, adult education, and income, and is used to model the difference between expected and actual values for all-cause mortality and a variety of other risk factors and diseases[5].

Selection Bias

To my knowledge, no study has been published that systematically evaluates the impact of selection bias on the estimation of fertility from birth histories; this thesis, however, draws inspiration from similar studies in child and adult mortality. In the case of child mortality, Coates[6] hypothesized that estimates of child mortality could be biased because the mortality of children of women who are not alive to be surveyed (orphans) could be systematically different from the mortality of non-orphans. By computing the ratio of child mortality between orphans and non-orphans (calculated from sibling histories) and the orphanhood prevalence (calculated from DHS household modules), he was able to calculate a correction factor for estimates of child mortality which ranged between near-0 to a 10% increase in child mortality in countries that were hit hard by the HIV epidemic, such as Zimbabwe. The United Nations Inter-agency Group for Child Mortality Estimation (IGME) also conducted analyses to correct for this bias using HIV prevalence and a model life table system[7].

Gakidou and King[8] proposed a correction for the impact of selection bias on estimates of adult mortality. Specifically, sibling histories suffer from selection bias because individuals from high mortality families are less likely to be included in the sample; by definition, families in which all siblings have died also are not included in the sample. The authors corrected this using the association between number of siblings in a sibship and the level of mortality found in DHS surveys.

Methods

Data

For this analysis, I made use of complete birth history estimates from major survey series including the Demographic and Health Surveys (DHS), the Multiple Indicator Cluster Survey (MICS), the Malaria Indicator Survey (MIS), the CDC Reproductive Health Survey (RHS), the World Fertility Survey (WFS) and the AIDS Indicator Survey (AIS), along with other country- or region-specific surveys, such as Arab League PAPCHILD/PAPFAM. A total of 299 surveys from 58 countries were used, spanning from a WFS conducted in Nepal in 1976 to two surveys conducted between 2016 and 2017, a MICS and a DHS conducted in Nigeria and Nepal respectively. The majority of survey data was taken from DHS, contributing 209 out of the 299 surveys analyzed. A complete list of sources is available in appendix table 9.

These surveys were obtained by taking the pool of all survey extracts with complete birth histories used in the fertility model in the Global Burden of Disease Project (GBD),[5] a pool of 468 surveys from 107 countries. Surveys or observations were then dropped according to the following criteria:

1. Surveys that were the only CBH for a given country, as my analysis requires at least two surveys. This represents 21 surveys from 21 countries.
2. Surveys in which required columns were missing, including mother id, survey date, mother date of birth, child id, and child date of birth. This represents 54 surveys from 29 countries.
3. Ever married surveys missing the all-women correction factor. These surveys were the ones from the late 1970s and represents 11 surveys from 11 locations.

4. Surveys from countries where criteria 2 or 3 resulted in having only one remaining CBH from the country.

Additionally, individual observations in included surveys were dropped if mother's date of birth or survey date was missing, or if a child id was present but the child date of birth was missing.

Calculation of Cohort Fertility

The basic principle of the first part of my analysis is that women surveyed in a single CBH survey provide information about their entire fertility history at every age previous. Thus, looking at a series of surveys, one can compute the ASFR of a cohort of women in a single age group from different surveys conducted at different points in time. In the absence of bias, one would expect these estimates to be very similar, differing only due to sampling error.

In order to calculate cohort fertility, I first calculated births by maternal age and year, excluding births and exposure in the month of the interview to prevent censoring, as DHS does, and then calculated the number of births per mother-age in case a given mother has more than one child born at a given age. Then, I calculated the exposure each mother contributes to each age in each calendar year. For example, if a mother is born in September, she contributes 9/12 years of exposure to one age and 3/12 years of exposure to the next age in a given year. In the year of the survey, the mother contributes the portion of her exposure in a given age group ending a month previous to the survey. For example, if a survey was conducted in February of a year and a mother's birthday was in October of the same year, she would contribute 1/12 years of exposure to the younger age group, and 0 years of exposure to the higher age group. Both exposure and births beyond a given recall period (I use 35 years) are also dropped.

Next, I calculated exposure and births for 5 year age cohorts from single-year exposure and births. That is, I took births and the person-time of the mother's exposure for a mother who is 15 years old in 1970, 16 years old in 1971, and up to 19 years old in 1974, and added the births and exposure per mother. In order to prevent censoring, I dropped observations for which a mother did not contribute all 5 years of exposure within each cohort. For example, if a mother was interviewed at age 33, I would keep her births and exposure for 5 year age groups up to 25-29, but drop births and exposure for the age group 30-34, since the mother didn't experience that entire exposure prior to the survey. This was to account for a potential bias due to the age pattern of fertility within a 5 year age group. For example, if fertility is decreasing between 35 and 39 years of age, not capturing the fertility behavior of 39 year olds would bias ASFR for women aged 35-39 upward.

Finally, to scale from births and exposure for a single mother to a nationally representative value, I applied the survey weights to account for unequal selection probabilities and the all-women factors in surveys that only sample ever-married women in order to account for the fact that not all women in the country are married.[4] Cohort ASFR was then obtained by dividing the resulting births by the resulting exposure. The final output of this analysis is the cohort ASFR for 5-year age groups for each survey included in the analysis. ASFR observations were dropped if they did not fall between 0.05 and 0.5 for 5 year age groups between 15-19 and 30-34. This led to dropping 923 observations between those age groups out of a total of 31,209 observations, or just under 3%.

Calculation of Adjustment Factor

In order to calculate an adjustment factor for potentially biased cohort ASFR calculations, I first kept observations with more than one estimate of cohort ASFR per cohort-age-country. I designated a reference ASFR as the cohort ASFR calculated from the survey with the lowest recall (i.e., the earliest survey for the given cohort, age, and location). The reference source was chosen this way to minimize the impact biases resulting from the passage of time. Using these observations, I created and fit a model of the log of the ratio between the reference ASFR and the survey-specific ASFR. For the earliest survey per cohort-age, this ratio empirically will always be 1.

I tested 18 models of this log ratio, with each model required to include the length of recall for each given survey, as well as the length of the recall for the reference source, in order to make predictions. Otherwise, the covariates present in the models include the age-specific maternal mortality ratio (for the period age group), the age-specific all-cause female mortality rate, 5 year age group, a random intercept on super region, and various interactions between the aforementioned variables (Appendix Tables 3-8).

To choose a model, I conducted 5-fold cross-validation by holding out random different groups of 20% of the data, resulting in each of the 18 models being fit 5 times on different subsets of 80% of the data, and then predicted on the 20% of holdouts to have a complete data set of observations predicted out of sample. I model fit statistics including r-squared, adjusted r-squared, and the root mean squared error (RMSE). These are presented in the results section in Table 1.

Results

In the cohort ASFRs computed directly from the surveys, the earliest birth cohort represented was the 1926 birth cohort from the 1967 WFS surveys in Mexico and Nepal. The latest cohort represented was the 1996 birth cohort from surveys conducted in 2016 in 12 countries, including Ethiopia, Nigeria, and Peru. The mean recall period in the sample was 13.7 years, while the median recall was 12.4 years; there were 7,299 observations (17%) with a recall of less than 5 years and 1,778 observations (4.2%) with a recall of greater than 30 years. The number of observations by age ranged from 9,112 for 15-19 year olds to 1,842 for 40-44 year olds.

Figure 1 shows a representative example of the results of the cohort ASFR calculation. Each colored line in the figure is a single birth cohort, with up to 10 cohorts per panel. Each point on the line represents a value of the ASFR for the cohort calculated from a different survey with a different recall period. The x-axis shows the year of the survey and also implies a recall value—as the year of the survey increases for a given cohort, the recall also increases. The results from this calculation generally show two things. First, the figures are suggestive of bias in that the cohort lines, which should be relatively flat (plus or minus sampling error), seem to be patterned or have a distinct directionality to their slope. Second, despite the observable trends, the estimates do show considerable sampling error through the fluctuations up and down between surveys. Other locations, ages, and cohort graphs can be found in the supplementary materials.

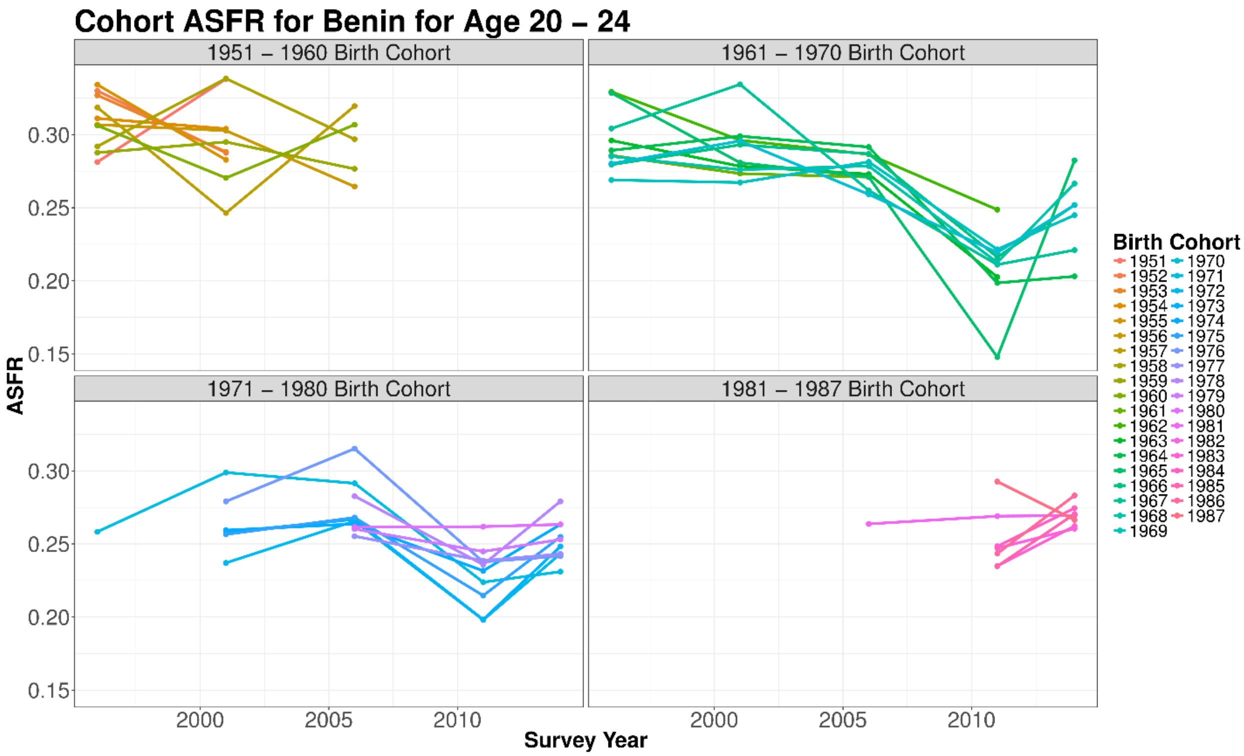


Figure 1

The mean of the reference recall, or the lowest recall for a given cohort-age within a location was 9.6 years, and the median was 8.00 years, with a max of 29 years, observed for the 1950 birth cohort in Cambodia. The empirical ratio of the reference ASFR to the survey-specific ASFR ranged between 0.21 and 5.86, signifying substantial sampling error to create such large differences. However, these empirical values were generally close to 1, with the mean at 1.02, the median at 1.00, and the interquartile range (IQR) between 0.94 and 1.05.

Eighteen models were fit on the log of the ratio between the reference and survey-specific ASFR, as described in the methods section. The results from the out of sample cross-validation are below in Table 1. Model specifications of the 18 models can be found in Tables 3 through 8 in the appendix. The highest adjusted R^2 value was found in model 9; however, the adjusted R^2

value for model 18 was similar, and the RMSE for model 18 was significantly lower, so I chose model 18.

Table 1: Model Fit Statistics

Model Number	R^2	Adjusted R^2	RMSE
1	0.013881	0.01379	0.006035
2	0.004351	0.004167	0.003872
3	0.014	0.013697	0.003834
4	0.018468	0.018242	0.006007
5	0.021983	0.021758	0.005986
6	0.014202	0.013717	0.003833
7	0.022352	0.021992	0.003802
8	0.026549	0.02628	0.005958
9	0.030398	0.03013	0.005934
10	0.022934	0.022394	0.003799
11	0.018292	0.018021	0.006008
12	0.013836	0.013472	0.003835
13	0.016192	0.015708	0.003826
14	0.019655	0.019173	0.003812
15	0.026512	0.026198	0.005958
16	0.023421	0.023001	0.003798
17	0.026305	0.025767	0.003786
18	0.0297	0.029163	0.003773

The specification of model 18 can be found in the equation below,

$$\log ratio = \beta_1 * recall + \beta_2 * reference\ recall + \beta_3 * age + \beta_4 * mx + \beta_5 * mmr + \beta_6 * mx * recall + \beta_7 * mmr * recall + \beta_8 * age * recall + \gamma_{sr}$$

where the reference recall refers to the recall from the reference ASFR for a given cohort-age-country, and γ_{sr} is a random intercept on GBD super region. This chosen model includes MMR, which is only modeled by the World Health Organization (WHO) from 1990 and by IHME from

1980, and therefore predictions were only made for birth-cohort-age pairings that were localized in time to 1980 or later. The coefficients of the chosen model are shown in Table 2 below, with the coefficients of the other tested models are shown in Tables 3 to 8 in the appendix.

Table 2: Best Model: Model 18

	Dependent variable:
	Log Ratio
recall	0.00479***
	(0.00086)
ref_recall	0.00162***
	(0.00044)
age	0.00200***
	(0.00050)
mx	3.83458***
	(1.01743)
mmr	−0.00016***
	(0.00001)
recall:mx	−0.53731***
	(0.12426)
recall:mmr	0.00002***
	(0.000001)
recall:age	−0.00038***
	(0.00004)
Constant	−0.02430*
	(0.01476)
Observations	20,368
Log Likelihood	3,450.34100
Akaike Inf. Crit.	−6,878.68200
Bayesian Inf. Crit.	−6,791.54300
Note:	*p<0.1; **p<0.05; ***p<0.01

When making predictions from this model, I set the reference recall to 0, that is, I predicted the ratio between reference and survey-specific ASFR as if the estimate was in the year of the survey. The modeled ratios ranged from a low of 0.66 for the 3 year recall in Timor-Leste for a birth cohort born in 1948 to a high of 1.40 for a 26 year recall also in Timor-Leste for a birth cohort born in 1965. The mean and median of the predicted ratios were both 0.98, meaning that on average, the survey specific ASFRs were biased upward in comparison to the reference ASFR. This finding was particularly strong in older ages, with the mean and median predicted ratios in the 35-39 year old age group being 0.96.

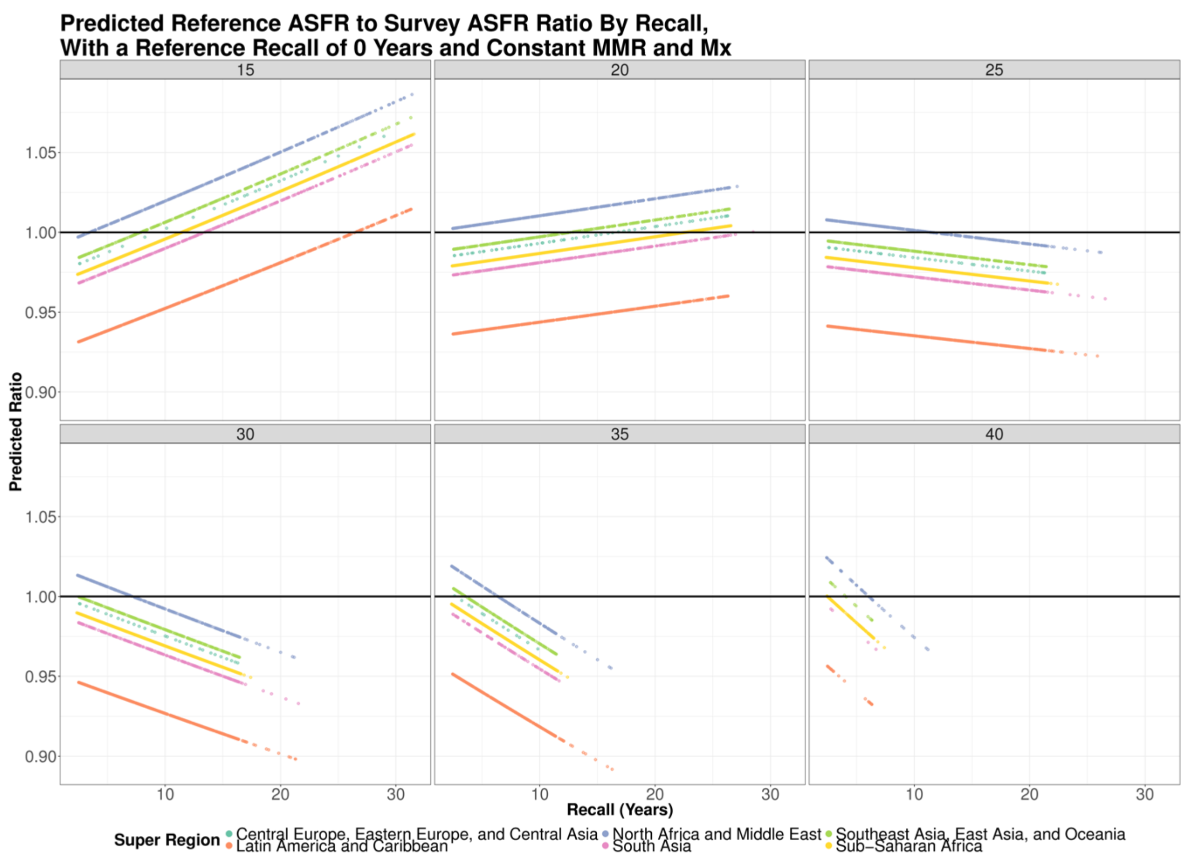
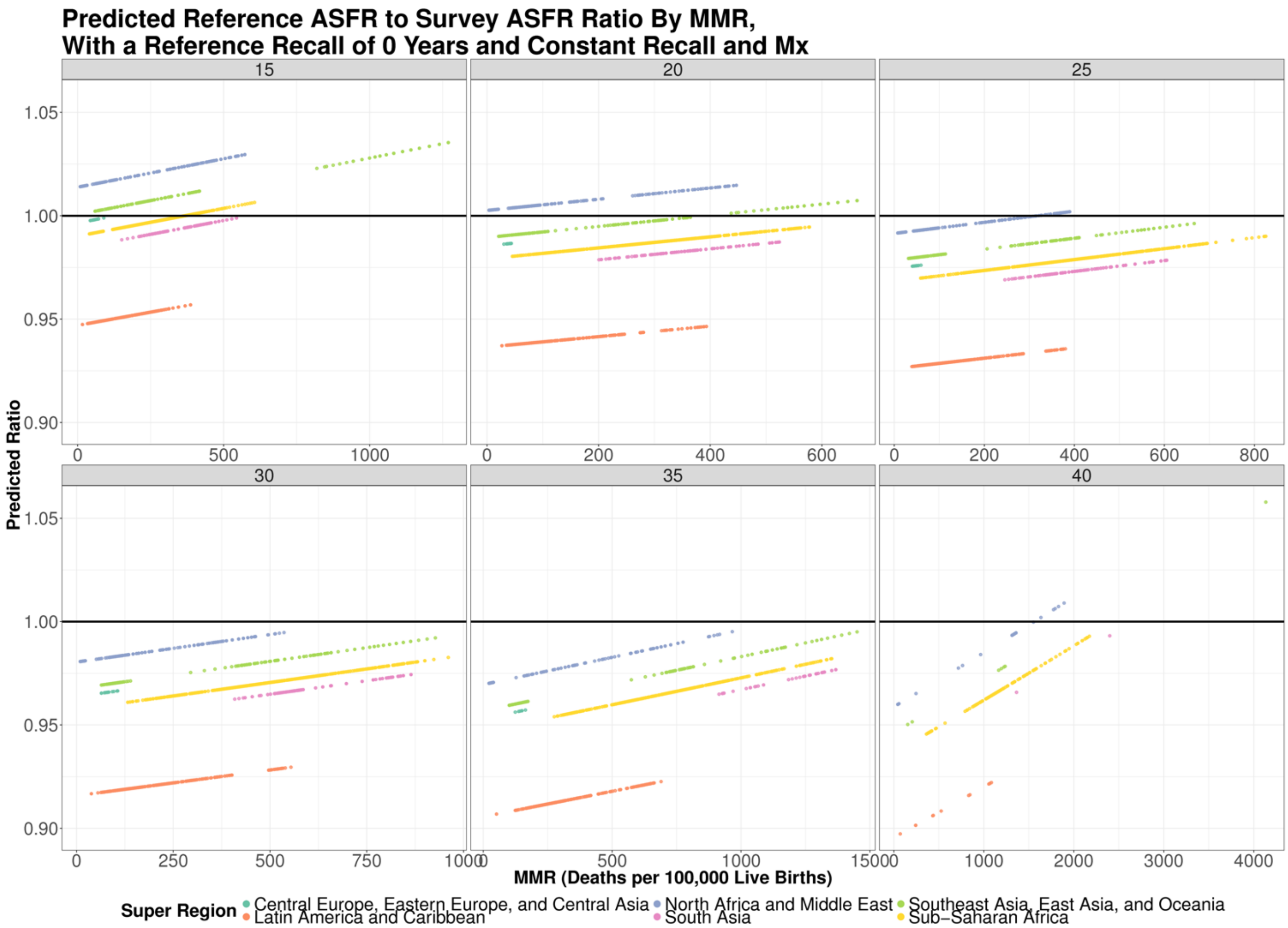


Figure 2

Figure 2 shows the relationship from the model between the predicted ratio and the number of years of recall, while MMR and adult female mortality were held constant at their mean. These results exhibit patterning by age, with ages 15-19 and 20-24 showing a positive relationship between the reference ASFR to survey-specific ASFR ratio and years of recall while ages 25-29, 30-34, 35-39, and 40-44 all show a negative relationship between the ratio and years of recall. These results also show that surveys in Latin America and the Caribbean seem to show significant over-estimation of fertility (especially in the older age groups) relative to the reference ASFR.

Figure 3 shows the relationship between the predicted ratio and MMR, with recall and adult female mortality kept constant at their averages, and figure 4 shows the relationship between the predicted ratio and adult female mortality with recall and MMR held constant at their averages. Figure 3 and figure 4 show that the relationships between the ratio and MMR and the ratio and female age specific mortality rate are in opposite directions, suggesting different underlying mechanisms.



Predicted Reference ASFR to Survey ASFR Ratio By MMR,
With a Reference Recall of 0 Years and Constant Recall and Mx
15
20
25
30
35
40
Predicted Ratio
MMR (Deaths per 100,000 Live Births)
Super Region
Central Europe, Eastern Europe, and Central Asia
North Africa and Middle East
Southeast Asia, East Asia, and Oceania
Latin America and Caribbean
South Asia
Sub–Saharan Africa

Figure 3

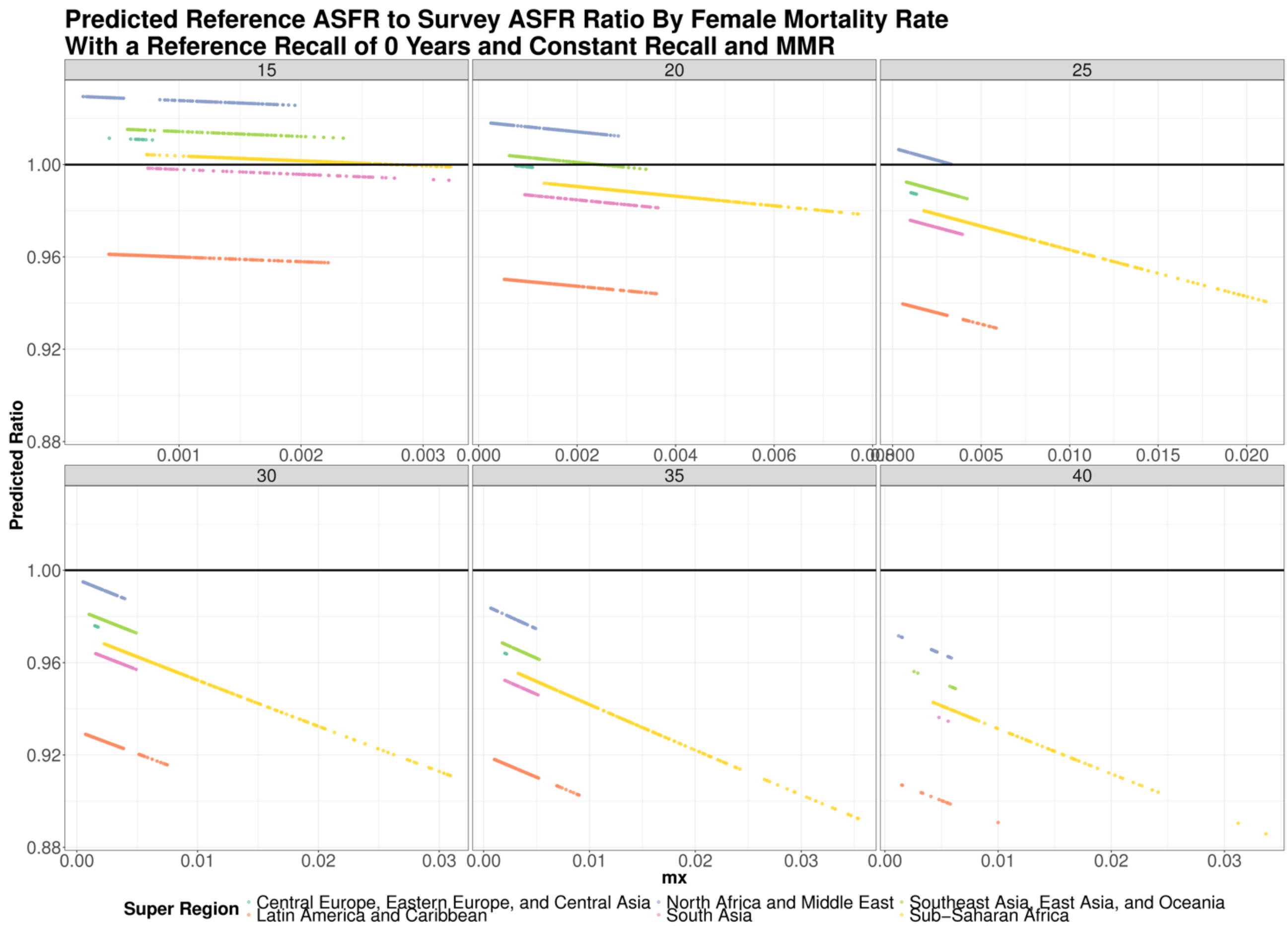


Figure 4

Figures 5 and 6 show the impact of applying the ratio as an adjustment factor by multiplying it by the survey-specific cohort ASFRs for two age-cohorts in Benin. The adjusted and unadjusted ASFRs are indicated by the labels, and the adjustment is in the direction that the triangle points. Different estimates for the same cohort are indicated by dotted lines connecting them.

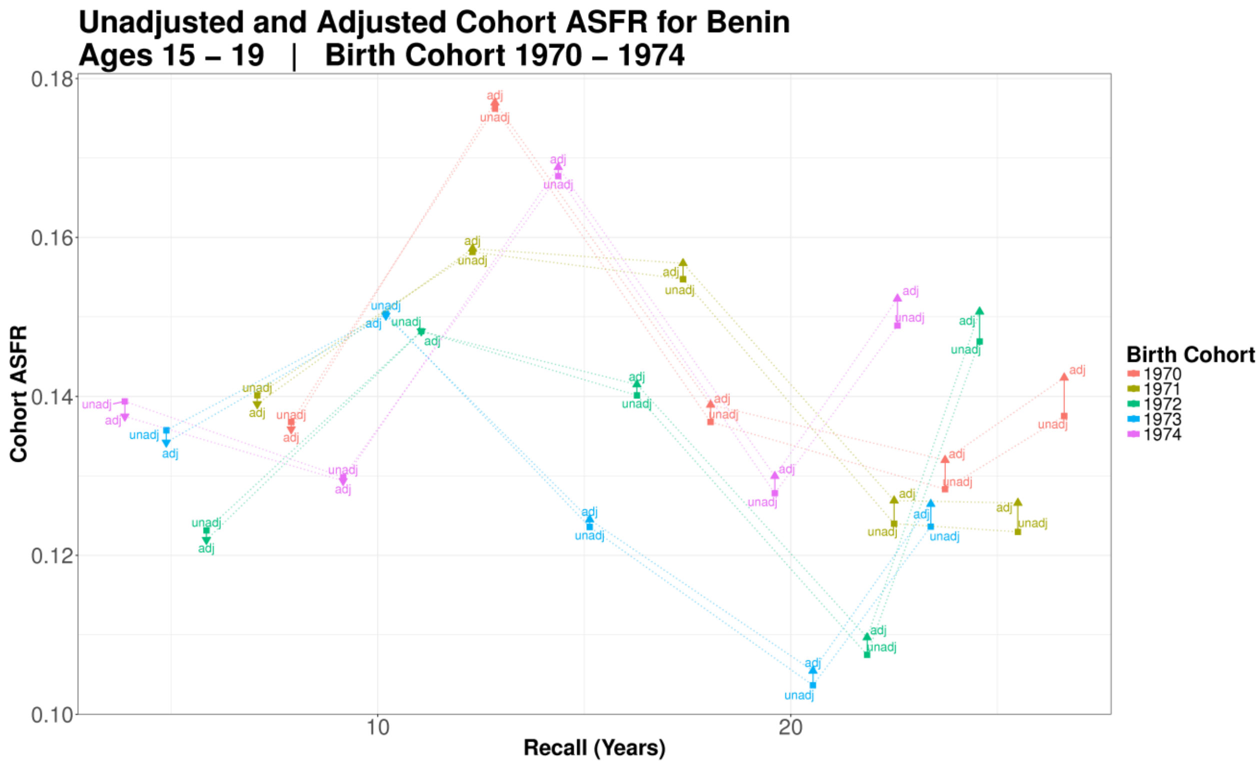


Figure 5

Figure 5 shows that for the Benin 1970-1974 cohort in ages 15-19, the directionality of the adjustment changes, with the data from surveys with lower recalls being adjusted down, the data from surveys with higher recalls being adjusted up, and the surveys between 8 and 15 years of recall being adjusted minimally. Figure 6 represents the same set of birth cohorts when they were aged 30-34. In this age group, all data is being adjusted downwards, with increasing recall being associated with increasing upward bias in the raw data points.

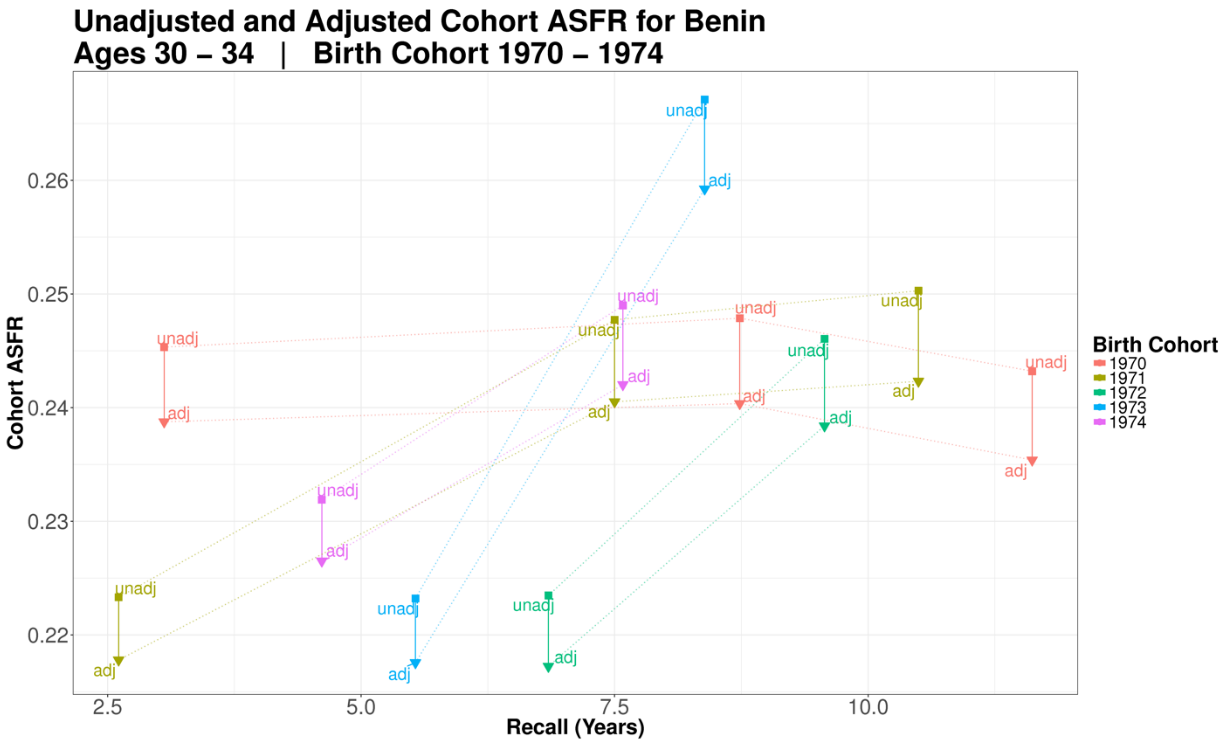


Figure 6

Discussion

My results show that the biases in estimates of ASFR for the same cohort-age differ based on the length of recall. For age group 15-19, increasing recall seems to be associated with increasing downward bias with recall; that is, as recall increases, estimates of ASFR for a given cohort decrease. On the other hand, for ages 25-29, 30-34, and 35-39 years, I observe the opposite relationship, with increasing recall being associated with higher estimates for a given cohort. These results are surprising, given the historical assumption in the literature that increasing recall leads to more forgetting of events —that assumption would mean a consistent relationship between recall and ASFR across ages.[9]

These relationships are suggestive of potential mechanisms behind the bias. These trends could be explained by selection bias, in which women who survive to be interviewed had systematically lower fertility when they were younger compared to those who died, and systematically higher fertility when they were older, also compared to those who died. For older ages, women who survive to be selected are more likely to have later births and thus higher older age ASFR; for younger ages, women who survive to be selected have maternal mortality as primary driver of mortality so have fewer children at a young age and thus lower young age ASFR.

The hypothesis that the primary driver of this bias is selection bias due to systematically different fertility among women who die and those who live seems to be supported by the opposite directionality of the relationship between the ratio and MMR and the ratio and adult female mortality as seen in figures 3 and 4. Another explanation for the direction in younger ages is the assumption that as recall increases, women forget their births at a higher rate.

Sources of Bias

There is agreement that there exists an association between time and accuracy of recall—the rates of forgetting being highest immediately following an event and then leveling off; however, there is also agreement that not all memories decay with time, and that salient events (such as giving birth) are more likely to be recalled.[9]

Most studies on bias in fertility estimation compare results from surveys to gold standard censuses, vital records, or demographic surveillance site (DSS) records; most only conduct analysis at the summary birth history level—that is, there are comparisons of the number of births per women, but these are not localized by age of women or in time.

In one study, the number of births reported by 786 women in the French West Indies in the 1954 census had high agreement when compared to the comprehensive vital records collected for these women.[10] While the author observed more errors in older women, age was not a predictor of either direction or magnitude of errors, suggesting that a straightforward bias due to "forgetting" was not present in this population, that is, systematic bias as recall/age at interview increases cannot be attributed to misremembering number of births, but rather to other factors.

Likewise, another study in rural Senegal found no evidence of under-reporting of births by age, even though their sample included women as old as 89; number of births was mostly consistent with the longitudinal DSS with only 2% fewer births reported in the survey.[11-12] The results from these studies are in line with my results in that they suggest that women misremembering their births is not the driving mechanism behind bias. The differing relationship between my adjustment factor and recall over age, as well as the result that at older ages, there is an increasing overestimation of fertility with increasing recall, also suggests other mechanisms for bias other than women forgetting their births.

On the other hand, Potter's scholarship[13] is discordant from my results, as he suggested that selection bias would have minimal impact on ASFR estimation as long as recall was restricted to 10 years and life expectancy exceeded 50. Potter was instead concerned with bias caused by a certain type of telescoping, in which women's recall in the very recent past is accurate, while they tend to attribute births in the distant past less distantly, leading to an underestimation of ASFR in the distant past and an overestimation in the middle period. Were this bias present in my results, one would expect the ratio in my model to be around 1, then decrease to under 1 and then increase to above 1 in the distant past. This trend does not seem to be present in the data or model in most regions. Moreover, in a later empirical study of the

Malaysia Family Life Surveys, authors found no evidence of telescoping in recall of children's birth dates.[14]

Two themes emerge from this discussion of past research on bias in fertility estimation. First, studies evaluating bias of fertility estimates due to the passage of time have been limited to single specific surveys in settings with accurate census, vital registration systems, or demographic surveillance sites. Surveys available to be analyzed this way are rare, because complete birth histories are generally not conducted in settings with the routine registration of vital events. Therefore the scope of previous work that is based on empirical data is limited, and results from these studies may not be completely generalizable given the unique circumstances allowing this kind of study.

Second, historically in the field of demography, there was often an assumption that if there existed systematic biases in collected survey data, it was due to women being surveyed forgetting the birth of their children or misremembering their children's date of birth. Demographers built formal demography models to address these issues, but Potter[13] was dismissive of other factors that could potentially cause bias, including, as Brittain[10] points out, survey question design and selection bias from differential mortality. Brittain further argued that, at least in the case of the French West Indies, the easiest way to measure fertility is simply to believe women about their births.

Importance

This thesis has attempted to describe, quantify, and correct for biases due to the passage of time in the estimation of cohort fertility from complete birth histories. To my knowledge, it represents the first systematic and empirical evaluation and correction of biases due to time in fertility estimation. Previous research on the bias in fertility estimation has either been attentive to very specific sources of bias,[3] based in formal demography and applied empirically,[13-14] or has focused on the evaluation of bias in single surveys.[9-12] Unfortunately, that means there are no comparable metrics against which to evaluate the performance of my correction factor.

One potential avenue of exploration in the future is the data adjustment step in IHME's fertility model,[5] in which a single source or series of sources are chosen as reference, and other sources within a given location are adjusted to that source using values from a random intercept on location-source. However, in the current implementation of this adjustment, entire time-series of sources are generally designated as reference, and each source gets the same adjustment regardless of recall. If only points from surveys within a certain limited recall period were designated as reference, and if each point previous got assigned a different random intercept, then the resulting adjustment to the data could serve as a comparison to the adjustment calculated by the methods in this analysis.

Using a correction similar to the one I have presented could allow IHME specifically to increase recall period used in the oldest age group of women surveyed from the current 3 years to 35 years, representing a large increase in the amount of fertility data back in time, especially in younger age groups. The importance of this additional data would be threefold. First, as fertility estimation currently extends back in time until 1950, it would theoretically allow for the inclusion of data in Sub-Saharan Africa from the first World Fertility Surveys to easily extend to that time period, potentially increasing the validity of estimates, as they would be based on corrected empirical data instead of on covariates and smoothing over space and time.

However, as the model currently includes MMR, which IHME only estimates back to 1980, this impact might be limited unless expanded estimation of MMR or the substitution of infant mortality rate for MMR, given their close relationship, is implemented. Alternately, some of the other models that performed well, including model 9, which had the highest adjusted R^2, do not include MMR in their equation. If model 9 passed sensitivity analyses, IHME could implement a model without MMR to adjust CBH fertility data with long recall periods back in time.

Second, such a correction factor could improve the prior estimates of fertility which currently do not vary in IHME's Bayesian population model, and thus improve estimates of population in the absence of census data or estimates of migration in contexts with good quality census data.

Third, total fertility under 25 (encompassing the age groups that would benefit most from increased recall period) is used in the estimation of IHME's Socio-Demographic Index (SDI) and thus in the modeling of expected results and the comparison of observed to expected results for morbidity, mortality, and risk factors. An increase in the data availability could increase the quality of these comparisons.

Limitations and Future Directions

This work is important because it offers a way to adjust any survey that includes a CBH module conducted anywhere to minimize recall, regardless of the exact mechanism behind the bias. However, my work has a number of limitations as well, which could be explored in the future.

First, the data themselves are somewhat unstable. I have tried to address this by collapsing to 5-year age cohorts, yet the estimates remain unstable and it can be difficult to disentangle true bias and trends in bias from the noise. Relatedly, I did not use the strata or primary sampling unit to calculate variance and get a measure of uncertainty around the estimates. If I did so, generated draws from the variance, and ran these draws through to the end of the model, I could get more precise information.

Second, this analysis did not explicitly address biases due to heaping and birth transference, and only addressed telescoping in a limited way.

Third, my entire analysis was done in cohort space. However, the estimates of MMR and adult female mortality that I pulled from the GBD 2016 study are both period measures, and IHME generally makes use of period measures.[5, 15] The future steps to address this limitation are twofold: first, use the lexis diagram and population[16] to calculate cohort estimates of MMR and adult female mortality to use in the model. Then, after applying the correction factor in cohort space, calculate period fertility adjustments using the same principles. I could also conduct a sensitivity analysis on the validity of going between cohort space and period space.

Fourth, some of the cohort ASFRs calculated from surveys were implausibly high, exceeding 0.5 for even lower fertility age groups. Because I included 299 surveys in my analysis to increase statistical power, I did not thoroughly investigate the data quality issues behind each instance of implausible ASFR and gain knowledge of each specific survey in my model; instead, I simply dropped those observations from my analysis.

Fifth, I decided to focus on complete birth history methods of estimating fertility. Demographers have long developed indirect methods of computing fertility from summary birth histories, and these also represent a substantial portion of the data used in the GBD ASFR model. However, the assumptions with indirect methods are much more stringent, and the interactions between those assumptions and survival bias are complex, putting these issues outside the scope of this project but presenting an interesting avenue to pursue in the future.

Conclusions

This study has shown that age-specific fertility rates are systematically biased as recall period increases back in time. The direction and magnitude of bias depend on the MMR, female adult age specific mortality rate, and the 5-year age group for which cohort ASFR is being estimated. An adjustment factor to correct estimates to a minimum recall period was proposed and implemented. This study is the first to systematically evaluate and correct bias in ASFR to the passage of time. More research is needed in order to validate the results in order for them to be incorporated into routine estimation of fertility.

References

1. WHO | SDG 3: Ensure healthy lives and promote wellbeing for all at all ages. WHO. http://www.who.int/sdg/targets/en/. Accessed June 8, 2018.

2. The DHS Program - DHS-Questionnaires. https://dhsprogram.com/What-We-Do/Survey-Types/DHS-Questionnaires.cfm. Accessed April 9, 2018.

3. Schoumaker B. Quality and consistency of DHS fertility estimates, 1990 to 2012. 2014. http://dhsprogram.com/publications/publication-MR12-Methodological-Reports.cfm. Accessed May 8, 2018.

4. Rutstein Oscar, Guillermo R. Guide to DHS Statistics. 2008; https://dhsprogram.com/publications/publication-dhsg1-dhs-questionnaires-and-manuals.cfm Accessed May 8, 2018.

5. Wang H, Abajobir AA, Abate KH, et al. Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1084-1150. doi:10.1016/S0140-6736(17)31833-0

6. Coates MM. Quantifying Selection Bias from Birth History Estimates of Child Mortality. Seattle]: University of Washington; 2016. http://hdl.handle.net/1773/36991. Accessed June 8, 2018.

7. Walker N, Hill K, Zhao F. Child Mortality Estimation: Methods Used to Adjust for Bias due to AIDS in Estimating Trends in Under-Five Mortality. PLoS Med. 2012;9(8). doi:10.1371/journal.pmed.1001298

8. Gakidou E, King G. Death by survey: Estimating adult mortality without selection bias from sibling survival data. Demography. 2006;43(3):569-585. doi:10.1353/dem.2006.0024

9. Beckett M, Da Vanzo J, Sastry N, Panis C, Peterson C. The Quality of Retrospective Data: An Examination of Long-Term Recall in a Developing Country. The Journal of Human Resources. 2001;36(3):593-625. doi:10.2307/3069631

10. Brittain AW. Can women remember how many children they have borne? Data from the east Caribbean. Social Biology. 1991;38(3-4):219-232. doi:10.1080/19485565.1991.9988789

11. Garenne M. Do women forget their births? A study of maternity histories in a rural area of Senegal (Niakhar). Popul Bull UN. 1994;(36):43-54.

12. Garenne M, Ginneken J van. Comparison of retrospective surveys with a longitudinal follow-up in Senegal: SFS, DHS and Niakhar. Eur J Population. 1994;10(3):203-221. doi:10.1007/BF01265302

13. Potter JE. Problems in using birth-history analysis to estimate trends in fertility. Population Studies. 1977;31(2):335-364. doi:10.1080/00324728.1977.10410433

14. United Nations. *Manual X: Indirect Techniques for Demographic Estimation*. United Nations publication. 1983; E.83.XIII.2

15. Naghavi M, Abajobir AA, Abbafati C, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1151-1210. doi:10.1016/S0140-6736(17)32152-9

16. Preston S, Heuveline P, Guillot M. Demography: Measuring and Modeling Population Processes. Blackwell Publishing; 2000.

Appendix

Table 3: Appendix Table

	Results from Models 1, 2, amd 3		
		Log Ratio	
	Model 1	Model 2	Model 3
recall	0.003***	0.001***	0.0001
	(0.0002)	(0.0002)	(0.0003)
ref_recall	0.001*	0.003***	0.002***
	(0.0003)	(0.0004)	(0.0004)
mmr		−0.00001	0.0003***
		(0.00001)	(0.00002)
age			−0.0002
			(0.0003)
mmr:age			−0.00001***
			(0.00000)
Constant	−0.044***	−0.040***	−0.047***
	(0.003)	(0.004)	(0.010)
Observations	27,195	20,368	20,368
R^2	0.014	0.005	0.015
Adjusted R^2	0.014	0.005	0.014
Residual Std. Error	0.206 (df = 27192)	0.206 (df = 20364)	0.205 (df = 20362)
F Statistic	195.456*** (df = 2; 27192)	31.790*** (df = 3; 20364)	60.748*** (df = 5; 20362)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: Appendix Table

	Results from Models 4, 5, and 6		
		Log Ratio	
	Model 4	Model 5	Model 6
recall	0.002***	0.008***	0.0001
	(0.0002)	(0.001)	(0.0003)
ref_recall	0.0002	−0.00004	0.002***
	(0.0003)	(0.0003)	(0.0004)
mmr			0.0003***
			(0.00003)
mx	6.434***		−7.558***
	(2.239)		(2.776)
mmr:age			−0.00001***
			(0.00000)
age:mx			0.261***
			(0.089)
age	−0.002***	0.001**	−0.0003
	(0.0003)	(0.0004)	(0.0004)
mx:age	−0.170**		
	(0.071)		
recall:age		−0.0003***	
		(0.00003)	
Constant	0.012	−0.040***	−0.041***
	(0.008)	(0.009)	(0.010)
Observations	27,195	27,195	20,368
R^2	0.019	0.022	0.015
Adjusted R^2	0.019	0.022	0.015
Residual Std. Error	0.205 (df = 27189)	0.205 (df = 27190)	0.205 (df = 20360)
F Statistic	105.774*** (df = 5; 27189)	156.135*** (df = 4; 27190)	44.741*** (df = 7; 20360)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5: Appendix Table

	Results from Models 7, 8, and 9		
		Log Ratio	
	Model 7	Model 8	Model 9
recall	0.0003	0.002***	0.009***
	(0.0003)	(0.0002)	(0.001)
ref_recall	0.001***	−0.0003	−0.001*
	(0.0004)	(0.0003)	(0.0003)
mmr	0.0002***		
	(0.00003)		
mx		4.664**	
		(2.376)	
age	−0.0003	−0.002***	0.001*
	(0.0003)	(0.0003)	(0.0004)
mmr:age	−0.00001***		
	(0.00000)		
mx:age		−0.125*	
		(0.073)	
recall:age			−0.0003***
			(0.00003)
Constant	−0.023*	0.025**	−0.030**
	(0.014)	(0.012)	(0.012)
Observations	20,368	27,195	27,195
Log Likelihood	3,381.465	4,587.680	4,632.171
Akaike Inf. Crit.	−6,746.929	−9,159.361	−9,250.342
Bayesian Inf. Crit.	−6,683.556	−9,093.674	−9,192.867
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6: Appendix Table

	Results from Models 10, 11, and 12		
	Log Ratio	Log Ratio	
	linear mixed-effects	*OLS*	
	Model 10	Model 11	Model 12
recall	0.0003	0.002***	−0.003***
	(0.0003)	(0.0003)	(0.0004)
ref_recall	0.001***	0.0002	0.003***
	(0.0004)	(0.0003)	(0.0004)
mmr	0.0002***		−0.0001***
	(0.00003)		(0.00001)
recall:mmr			0.00001***
			(0.00000)
age	−0.0004	−0.003***	−0.002***
	(0.0004)	(0.0002)	(0.0003)
mx	−11.349***	1.593**	
	(2.928)	(0.782)	
mmr:age	−0.00001***		
	(0.00000)		
age:mx	0.364***		
	(0.092)		
recall:mx		−0.056	
		(0.087)	
Constant	−0.015	0.021***	0.039***
	(0.014)	(0.007)	(0.008)
Observations	20,368	27,195	20,368
R^2		0.019	0.015
Adjusted R^2		0.019	0.014
Log Likelihood	3,388.070		
Akaike Inf. Crit.	−6,756.140		
Bayesian Inf. Crit.	−6,676.923		
Residual Std. Error		0.205 (df = 27189)	0.205 (df = 20362)
F Statistic		104.690*** (df = 5; 27189)	60.111*** (df = 5; 20362)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7: Appendix Table

	Results from Models 13, 14, and 15		
	Log Ratio		Log Ratio
	OLS		*linear mixed-effects*
	Model 13	Model 14	Model 15
recall	−0.002***	0.004***	0.003***
	(0.0004)	(0.001)	(0.0003)
ref_recall	0.003***	0.003***	−0.0002
	(0.0004)	(0.0004)	(0.0003)
age	−0.002***	0.001***	−0.002***
	(0.0003)	(0.0005)	(0.0002)
mx	6.543***	3.764***	1.499*
	(0.971)	(1.021)	(0.786)
mmr	−0.0001***	−0.0001***	
	(0.00001)	(0.00001)	
recall:mx	−0.805***	−0.431***	−0.113
	(0.115)	(0.123)	(0.089)
recall:mmr	0.00002***	0.00002***	
	(0.00000)	(0.00000)	
recall:age		−0.0004***	
		(0.00004)	
Constant	0.031***	−0.033***	0.029***
	(0.008)	(0.011)	(0.011)
Observations	20,368	20,368	27,195
R^2	0.017	0.021	
Adjusted R^2	0.017	0.020	
Log Likelihood			4,587.235
Akaike Inf. Crit.			−9,158.469
Bayesian Inf. Crit.			−9,092.783
Residual Std. Error	0.205 (df = 20360)	0.205 (df = 20359)	
F Statistic	50.283*** (df = 7; 20360)	53.493*** (df = 8; 20359)	

Note: *p<0.1; **p<0.05; ***p<0.01

Table 8: Appendix Table

	Results from Models 16, 17, and 18		
		Log Ratio	
	Model 16	Model 17	Model 18
recall	−0.002***	−0.002***	0.005***
	(0.0004)	(0.0004)	(0.001)
ref_recall	0.002***	0.001***	0.002***
	(0.0004)	(0.0004)	(0.0004)
age	−0.002***	−0.002***	0.002***
	(0.0003)	(0.0003)	(0.0005)
mx		6.558***	3.835***
		(0.969)	(1.017)
mmr	−0.0001***	−0.0001***	−0.0002***
	(0.00001)	(0.00001)	(0.00001)
recall:mx		−0.918***	−0.537***
		(0.116)	(0.124)
recall:mmr	0.00001***	0.00002***	0.00002***
	(0.00000)	(0.00000)	(0.00000)
recall:age			−0.0004***
			(0.00004)
Constant	0.050***	0.039***	−0.024*
	(0.013)	(0.013)	(0.015)
Observations	20,368	20,368	20,368
Log Likelihood	3,392.200	3,422.302	3,450.341
Akaike Inf. Crit.	−6,768.399	−6,824.603	−6,878.682
Bayesian Inf. Crit.	−6,705.025	−6,745.386	−6,791.543

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9: Appendix Table: Survey Sources Used

NIDs are unique identifiers by survey that can be found on the GHDx.

Country	Source	Survey Start	Survey End	nid
Algeria	PAPFAM	2002	2002	627
Algeria	Unicef Multiple Indicator Cluster Survey	2012	2013	210614
Angola	Malaria Indicator Survey	2007	2007	672
Angola	Malaria Indicator Survey	2011	2011	56169
Angola	Demographic and Health Survey	2015	2016	218555
Bangladesh	Demographic and Health Survey	1993	1994	18889
Bangladesh	Demographic and Health Survey	1996	1997	18878
Bangladesh	Demographic and Health Survey	1999	2000	26826
Bangladesh	Demographic and Health Survey	2001	2001	18920
Bangladesh	Demographic and Health Survey	2004	2004	18902
Bangladesh	Bangladesh Urban Health Survey	2006	2006	95474
Bangladesh	Demographic and Health Survey	2007	2007	18913
Bangladesh	Demographic and Health Survey	2011	2012	55956
Bangladesh	Demographic and Health Survey	2014	2014	157021
Benin	Demographic and Health Survey	1996	1996	18938
Benin	Demographic and Health Survey	2001	2001	18950
Benin	Demographic and Health Survey	2006	2006	18959
Benin	Demographic and Health Survey	2011	2012	79839
Benin	Unicef Multiple Indicator Cluster Survey	2014	2014	206075
Bolivia	Demographic and Health Survey	1989	1989	18979
Bolivia	Demographic and Health Survey	1993	1994	18990
Bolivia	Demographic and Health Survey	1998	1998	18971
Bolivia	Demographic and Health Survey	2003	2004	19001
Bolivia	Demographic and Health Survey	2008	2008	19016
Brazil	Demographic and Health Survey	1986	1986	19027
Brazil	Demographic and Health Survey	1991	1991	19035
Brazil	Demographic and Health Survey	1996	1996	19046
Burkina Faso	Demographic and Health Survey	1992	1993	19064
Burkina Faso	Demographic and Health Survey	1998	1999	19076
Burkina Faso	Demographic and Health Survey	2003	2003	19088
Burkina Faso	Demographic and Health Survey	2010	2011	19133
Cambodia	Demographic and Health Survey	1998	1998	19170
Cambodia	Demographic and Health Survey	2000	2000	19156
Cambodia	Demographic and Health Survey	2005	2006	19167
Cambodia	Demographic and Health Survey	2010	2011	30379
Cambodia	Demographic and Health Survey	2014	2014	157024
Cameroon	Demographic and Health Survey	1991	1991	19188

Cameroon	Demographic and Health Survey	1998	1998	19198
Cameroon	Demographic and Health Survey	2004	2004	19211
Cameroon	Unicef Multiple Indicator Cluster Survey	2006	2006	2063
Cameroon	Demographic and Health Survey	2011	2011	19274
Cameroon	Unicef Multiple Indicator Cluster Survey	2014	2014	244455
Colombia	Demographic and Health Survey	1986	1986	19333
Colombia	Demographic and Health Survey	1990	1990	19341
Colombia	Demographic and Health Survey	1995	1995	19350
Colombia	Demographic and Health Survey	2000	2000	19359
Colombia	Demographic and Health Survey	2004	2005	19324
Colombia	Demographic and Health Survey	2009	2010	21281
Colombia	Demographic and Health Survey	2015	2016	218566
Dominican Republic	World Fertility Survey	1980	1980	3499
Dominican Republic	Demographic and Health Survey	1986	1986	19400
Dominican Republic	Demographic and Health Survey	1991	1991	19410
Dominican Republic	Demographic and Health Survey	1996	1996	19421
Dominican Republic	Demographic and Health Survey	1999	1999	19431
Dominican Republic	Demographic and Health Survey	2002	2002	19444
Dominican Republic	Demographic and Health Survey	2007	2007	19456
Dominican Republic	Demographic and Health Survey	2007	2007	21198
Dominican Republic	Demographic and Health Survey	2013	2013	77819
Dominican Republic	Unicef Multiple Indicator Cluster Survey	2014	2014	200697
Ecuador	World Fertility Survey	1979	1980	3558
Ecuador	Demographic and Health Survey	1987	1987	19464
Ecuador	CDC Reproductive Health Survey	1994	1994	27615
Ecuador	CDC Reproductive Health Survey	1999	1999	27621
Ecuador	CDC Reproductive Health Survey	2004	2004	27630
Egypt	World Fertility Survey	1980	1980	3603
Egypt	Demographic and Health Survey	1988	1989	19472
Egypt	Demographic and Health Survey	1992	1993	19482
Egypt	Demographic and Health Survey	1995	1996	19493
Egypt	Demographic and Health Survey	2000	2000	19511
Egypt	Demographic and Health Survey	2003	2003	19529
Egypt	Demographic and Health Survey	2005	2005	19521

Egypt	Demographic and Health Survey	2008	2008	26842
Egypt	Demographic and Health Survey	2014	2014	154897
El Salvador	Demographic and Health Survey	1985	1985	19533
El Salvador	CDC Reproductive Health Survey	1993	1993	27582
El Salvador	CDC Reproductive Health Survey	1998	1998	27590
El Salvador	CDC Reproductive Health Survey	2002	2003	27599
El Salvador	Unicef Multiple Indicator Cluster Survey	2014	2014	200636
Ethiopia	Demographic and Health Survey	2000	2000	19571
Ethiopia	Demographic and Health Survey	2005	2005	19557
Ethiopia	Demographic and Health Survey	2010	2011	21301
Ethiopia	Demographic and Health Survey	2016	2016	218568
Ghana	World Fertility Survey	1979	1980	26939
Ghana	Demographic and Health Survey	1988	1988	19587
Ghana	Demographic and Health Survey	1993	1994	19604
Ghana	Demographic and Health Survey	2003	2003	19627
Ghana	Demographic and Health Survey	1998	1999	19614
Ghana	Demographic and Health Survey	2007	2008	21173
Ghana	Unicef Multiple Indicator Cluster Survey	2007	2008	160576
Ghana	Demographic and Health Survey	2008	2008	21188
Ghana	Unicef Multiple Indicator Cluster Survey	2011	2011	63993
Ghana	Demographic and Health Survey	2014	2014	157027
Guatemala	Demographic and Health Survey	1987	1987	19647
Guatemala	Demographic and Health Survey	1995	1995	19637
Guatemala	Demographic and Health Survey	1998	1999	19656
Guatemala	CDC Reproductive Health Survey	2002	2002	27563
Guatemala	CDC Reproductive Health Survey	2008	2009	4779
Guatemala	Demographic and Health Survey	2014	2015	157031
Guinea	Demographic and Health Survey	1999	1999	19670
Guinea	Demographic and Health Survey	2005	2005	19683
Guinea	Demographic and Health Survey	2012	2012	69761
Guyana	AIDS Indicator Survey	2005	2005	4837
Guyana	Demographic and Health Survey	2009	2009	21348
Guyana	Unicef Multiple Indicator Cluster Survey	2014	2014	200598
Haiti	Demographic and Health Survey	1994	1995	19695
Haiti	Demographic and Health Survey	2000	2000	19708
Haiti	Demographic and Health Survey	2005	2006	19720
Haiti	Demographic and Health Survey	2012	2012	65118
Honduras	CDC Reproductive Health Survey	1996	1996	27542
Honduras	CDC Reproductive Health Survey	2001	2001	27551
Honduras	Demographic and Health Survey	2005	2006	19728
Honduras	Demographic and Health Survey	2011	2012	95440

India	Demographic and Health Survey	1992	1993	19787
India	India Human Development Survey	2011	2012	165498
India	Demographic and Health Survey	2005	2006	19963
India	Demographic and Health Survey	2015	2016	157050
Indonesia	Demographic and Health Survey	1987	1987	19970
Indonesia	Demographic and Health Survey	1991	1991	19979
Indonesia	Demographic and Health Survey	1994	1994	19990
Indonesia	Demographic and Health Survey	1997	1997	19999
Indonesia	Demographic and Health Survey	2002	2003	20011
Indonesia	Demographic and Health Survey	2007	2007	20021
Indonesia	Demographic and Health Survey	2012	2012	76705
Iraq	Unicef Multiple Indicator Cluster Survey	2006	2006	7028
Iraq	Unicef Multiple Indicator Cluster Survey	2011	2011	76707
Kyrgyzstan	Demographic and Health Survey	1997	1997	20154
Kyrgyzstan	Demographic and Health Survey	2012	2012	77518
Kyrgyzstan	Unicef Multiple Indicator Cluster Survey	2014	2014	162283
Lesotho	Demographic and Health Survey	2004	2005	20167
Lesotho	Demographic and Health Survey	2009	2010	21382
Lesotho	Demographic and Health Survey	2014	2014	157058
Liberia	Demographic and Health Survey	1986	1986	20180
Liberia	Demographic and Health Survey	2006	2007	20191
Liberia	Malaria Indicator Survey	2008	2009	34279
Liberia	Demographic and Health Survey	2013	2013	77385
Madagascar	Demographic and Health Survey	1992	1992	20202
Madagascar	Demographic and Health Survey	1997	1997	20212
Madagascar	Demographic and Health Survey	2003	2004	20223
Madagascar	Demographic and Health Survey	2008	2009	21409
Malawi	Demographic and Health Survey	1992	1992	20235
Malawi	Demographic and Health Survey	2000	2000	20252
Malawi	Demographic and Health Survey	2004	2005	20263
Malawi	Unicef Multiple Indicator Cluster Survey	2006	2006	7919
Malawi	Demographic and Health Survey	2010	2010	21393
Malawi	Unicef Multiple Indicator Cluster Survey	2013	2014	161662
Malawi	Demographic and Health Survey	2015	2016	218581
Mali	Demographic and Health Survey	1987	1987	20283
Mali	Demographic and Health Survey	1995	1996	20301
Mali	Demographic and Health Survey	2001	2001	20315
Mali	Demographic and Health Survey	2006	2006	20274
Mali	Demographic and Health Survey	2012	2013	77388
Mali	Unicef Multiple Indicator Cluster Survey	2015	2015	248224
Mali	Malaria Indicator Survey	2015	2015	218587

Mauritania	Demographic and Health Survey	2000	2001	20322
Mauritania	Unicef Multiple Indicator Cluster Survey	2011	2011	152783
Mexico	World Fertility Survey	1976	1977	8651
Mexico	Demographic and Health Survey	1987	1987	20326
Mexico	Unicef Multiple Indicator Cluster Survey	2006	2006	23982
Mexico	Unicef Multiple Indicator Cluster Survey	2009	2009	24006
Morocco	World Fertility Survey	1980	1980	8867
Morocco	Demographic and Health Survey	1987	1987	20347
Morocco	Demographic and Health Survey	1992	1992	20371
Morocco	Demographic and Health Survey	1995	1995	20351
Morocco	Demographic and Health Survey	2003	2004	20361
Mozambique	Demographic and Health Survey	1997	1997	20382
Mozambique	Demographic and Health Survey	2003	2004	20394
Mozambique	Unicef Multiple Indicator Cluster Survey	2008	2009	27031
Mozambique	Demographic and Health Survey	2011	2011	55975
Namibia	Demographic and Health Survey	1992	1992	20404
Namibia	Demographic and Health Survey	2000	2000	20417
Namibia	Demographic and Health Survey	2006	2007	20428
Namibia	Demographic and Health Survey	2013	2013	150382
Nepal	World Fertility Survey	1976	1976	9230
Nepal	Demographic and Health Survey	1996	1996	20437
Nepal	Demographic and Health Survey	2001	2001	20450
Nepal	Demographic and Health Survey	2006	2006	20462
Nepal	Demographic and Health Survey	2011	2011	21240
Nepal	Demographic and Health Survey	2016	2017	286782
Nicaragua	CDC Reproductive Health Survey	1992	1993	9278
Nicaragua	Demographic and Health Survey	1997	1998	20478
Nicaragua	Demographic and Health Survey	2001	2001	20487
Nicaragua	CDC Reproductive Health Survey	2006	2007	9270
Nicaragua	Demographic and Health Survey	2011	2012	126952
Niger	Demographic and Health Survey	1992	1992	20518
Niger	Demographic and Health Survey	1998	1998	20537
Niger	Demographic and Health Survey	2006	2006	20499
Niger	Demographic and Health Survey	2012	2012	74393
Nigeria	Malaria Indicator Survey	2010	2010	30991
Nigeria	Unicef Multiple Indicator Cluster Survey	2016	2017	218613
Nigeria	Malaria Indicator Survey	2015	2015	218590
Palestine	PAPFAM	2006	2007	9999
Palestine	Unicef Multiple Indicator Cluster Survey	2010	2010	125591
Palestine	Unicef Multiple Indicator Cluster Survey	2014	2014	161590
Paraguay	World Fertility Survey	1979	1979	10400

Paraguay	Demographic and Health Survey	1990	1990	20608
Paraguay	CDC Reproductive Health Survey	1995	1996	10364
Paraguay	CDC Reproductive Health Survey	2004	2004	10370
Paraguay	CDC Reproductive Health Survey	2008	2008	27525
Paraguay	Unicef Multiple Indicator Cluster Survey	2016	2016	324470
Peru	World Fertility Survey	1977	1978	10890
Peru	Demographic and Health Survey	1986	1986	20617
Peru	Demographic and Health Survey	1991	1992	20626
Peru	Demographic and Health Survey	1996	1996	20638
Peru	Demographic and Health Survey	2000	2000	20649
Peru	Demographic and Health Survey	2003	2008	275090
Peru	Demographic and Health Survey	2009	2009	270404
Peru	Demographic and Health Survey	2010	2010	270469
Peru	Demographic and Health Survey	2011	2011	270470
Peru	Demographic and Health Survey	2012	2012	270471
Peru	Demographic and Health Survey	2013	2013	146860
Peru	Demographic and Health Survey	2014	2014	209930
Peru	Peru Demographic and Family Health Survey	2015	2015	303663
Peru	Peru Demographic and Family Health Survey	2016	2016	303664
Philippines	Demographic and Health Survey	1993	1993	20674
Philippines	Demographic and Health Survey	1998	1998	20683
Philippines	Demographic and Health Survey	2003	2003	20699
Philippines	Demographic and Health Survey	2008	2008	21421
Philippines	Demographic and Health Survey	2011	2011	135803
Philippines	Demographic and Health Survey	2013	2013	142943
Rwanda	World Fertility Survey	1983	1983	11348
Rwanda	Demographic and Health Survey	1992	1992	20711
Rwanda	Demographic and Health Survey	2000	2000	20722
Rwanda	Demographic and Health Survey	2005	2005	20740
Rwanda	Demographic and Health Survey	2007	2008	21222
Rwanda	Demographic and Health Survey	2010	2011	56040
Rwanda	Demographic and Health Survey	2014	2015	157063
Senegal	Demographic and Health Survey	1986	1986	20749
Senegal	Demographic and Health Survey	1992	1993	20767
Senegal	Demographic and Health Survey	1997	1997	20780
Senegal	Demographic and Health Survey	1999	1999	20786
Senegal	Demographic and Health Survey	2005	2005	26855
Senegal	Malaria Indicator Survey	2006	2006	11516
Senegal	Malaria Indicator Survey	2009	2009	11540
Senegal	Demographic and Health Survey	2010	2011	56063
Senegal	Demographic and Health Survey	2012	2013	111432

Senegal	Demographic and Health Survey	2014	2014	191270
Senegal	Demographic and Health Survey	2015	2015	218592
Senegal	Demographic and Health Survey	2016	2016	286772
Sierra Leone	Demographic and Health Survey	2008	2008	21258
Sierra Leone	Demographic and Health Survey	2013	2013	131467
Sierra Leone	Malaria Indicator Survey	2016	2016	286773
Sudan	Demographic and Health Survey	1989	1990	20813
Sudan	PAPCHILD	1992	1993	12259
Sudan	Unicef Multiple Indicator Cluster Survey	2010	2010	32189
Sudan	Unicef Multiple Indicator Cluster Survey	2010	2010	153643
Sudan	Unicef Multiple Indicator Cluster Survey	2014	2014	200617
Swaziland	Demographic and Health Survey	2006	2007	20829
Swaziland	Unicef Multiple Indicator Cluster Survey	2010	2010	30325
Swaziland	Unicef Multiple Indicator Cluster Survey	2014	2014	200707
Tanzania	Demographic and Health Survey	1991	1992	20841
Tanzania	Demographic and Health Survey	1996	1996	20852
Tanzania	Demographic and Health Survey	1999	1999	20865
Tanzania	Demographic and Health Survey	2004	2005	20875
Tanzania	Demographic and Health Survey	2009	2010	21331
Tanzania	AIDS Indicator Survey	2011	2012	77395
Tanzania	Demographic and Health Survey	2015	2016	218593
Timor-Leste	Demographic and Health Survey	1994	1994	19990
Timor-Leste	Demographic and Health Survey	1997	1997	19999
Timor-Leste	Demographic and Health Survey	2003	2003	20888
Timor-Leste	Demographic and Health Survey	2009	2010	21274
Togo	Demographic and Health Survey	1988	1988	20896
Togo	Demographic and Health Survey	1998	1998	20909
Togo	Unicef Multiple Indicator Cluster Survey	2000	2000	12886
Togo	Demographic and Health Survey	2013	2014	77515
Tunisia	World Fertility Survey	1978	1978	26957
Tunisia	Demographic and Health Survey	1988	1988	20926
Tunisia	PAPFAM	2001	2001	12978
Tunisia	Unicef Multiple Indicator Cluster Survey	2011	2012	76709
Turkey	World Fertility Survey	1978	1978	13058
Turkey	Demographic and Health Survey	1993	1993	20936
Turkey	Demographic and Health Survey	1998	1998	20947
Turkey	Demographic and Health Survey	2003	2004	20954
Turkey	Demographic and Health Survey	2008	2008	32421
Uganda	Demographic and Health Survey	1988	1989	20964
Uganda	Demographic and Health Survey	1995	1995	20976
Uganda	Demographic and Health Survey	2000	2001	20993

Uganda	Demographic and Health Survey	2006	2006	21014
Uganda	Malaria Indicator Survey	2009	2010	13109
Uganda	Demographic and Health Survey	2011	2011	56021
Uganda	Demographic and Health Survey	2016	2016	286780
Vietnam	Demographic and Health Survey	1997	1997	21049
Vietnam	Demographic and Health Survey	2002	2002	21058
Vietnam	Unicef Multiple Indicator Cluster Survey	2013	2014	152735
Yemen	Demographic and Health Survey	1991	1992	21068
Yemen	Unicef Multiple Indicator Cluster Survey	2006	2006	13816
Yemen	Demographic and Health Survey	2013	2013	112500
Zambia	Demographic and Health Survey	1992	1992	21079
Zambia	Demographic and Health Survey	1996	1997	21090
Zambia	Demographic and Health Survey	2001	2002	21102
Zambia	Demographic and Health Survey	2007	2007	21117
Zambia	Demographic and Health Survey	2013	2014	77516
Zimbabwe	Demographic and Health Survey	1988	1989	21126
Zimbabwe	Demographic and Health Survey	1994	1994	21139
Zimbabwe	Demographic and Health Survey	1999	1999	21151
Zimbabwe	Demographic and Health Survey	2005	2006	21163
Zimbabwe	Unicef Multiple Indicator Cluster Survey	2009	2009	35493
Zimbabwe	Demographic and Health Survey	2010	2011	55992
Zimbabwe	Unicef Multiple Indicator Cluster Survey	2014	2014	152720
Zimbabwe	Demographic and Health Survey	2015	2015	157066