USAID
FROM THE AMERICAN PEOPLE

AN ASSESSMENT OF THE QUALITY AND CONSISTENCY OF AGE AND DATE REPORTING IN DHS SURVEYS, 2000-2015

DHS METHODOLOGICAL REPORTS 19

AUGUST 2017

This publication was produced for review by the United States Agency for International Development. It was prepared by Thomas Pullum and Sarah Staveteig.

DHS Methodological Reports No. 19

An Assessment of the Quality and Consistency of Age and Date Reporting in DHS Surveys, 2000-2015

Thomas Pullum[1,2,§]
Sarah Staveteig[2,3,§]

ICF

Rockville, Maryland, USA

August 2017

[1] ICF
[2] The DHS Program
[3] Avenir Health

[§] Both authors contributed equally to this work

Corresponding author: Thomas Pullum, International Health and Development, ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA; phone: 301–572–0950; fax: 301–572–0999; email: tom.pullum@icf.com

Acknowledgments: The authors wish to thank Stan Becker for valuable comments, and Fred Arnold, Trevor Croft, and Anjushree Pradhan for helpful discussions of DHS fieldwork procedures.

Editor: Bryant Robey

Document Production: Christopher Gramer

This study was carried out with support provided by the United States Agency for International Development (USAID) through The DHS Program (#AIDOAA-C-13-00095). The views expressed are those of the authors and do not necessarily reflect the views of USAID or the United States Government.

The DHS Program assists countries worldwide in the collection and use of data to monitor and evaluate population, health, and nutrition programs. For additional information about The DHS Program contact: DHS Program, ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA. Phone: +1 301-407-6500; fax: +1 301-407-6501; email: reports@dhsprogram.com; Internet: www.dhsprogram.com.

Recommended citation:

Pullum, Thomas, and Sarah Staveteig. 2017. *An Assessment of the Quality and Consistency of Age and Date Reporting in DHS Surveys, 2000-2015*. DHS Methodological Reports No. 19. Rockville, Maryland, USA: ICF.

Contents

Tables ..... **v**
Figures ..... **vii**
Preface ..... **ix**
Abstract ..... **xi**

1 Background **1**
1.1 DHS Age and Date Questions Analyzed in this Report 1
1.1.1 Household Questionnaire 1
1.1.2 Woman's Questionnaire 1
1.1.3 Man's Questionnaire 2
1.2 The DHS Program's Procedures for Obtaining Accurate Age and Date Information 2
1.3 The DHS Program's Procedure for Editing Inconsistent and Imputing Incomplete Dates 3
1.4 Prior DHS Studies on Data Quality 4
1.5 Report Structure 4

2 Incompleteness, Heaping, and Displacement in DHS Surveys **7**
2.1 Data and Methods 7
2.2 Incompleteness 10
2.3 Heaping 14
2.4 Displacement 18
2.5 Composite Measures of Data Quality 23

3 The Quality and Consistency of Spousal Age Estimates in DHS Surveys **27**
3.1 Data and Methods 27
3.2 Composition of Couples by Direction of Age Difference 28
3.3 Difference between Spousal Age Estimates and Self-Reports 28
3.4 Heaping in Spousal Age Estimates 31
3.5 Consistency of Spousal Estimates by Age 35
3.6 Consistency of Spousal Age Estimates by Education 36

4 Conclusion **41**

References **47**

Appendix A: All Ages, Dates, and Durations Asked of DHS Respondents and their Evolution between DHS Phases 4 to 7 ..... **49**
Household and Biomarker Questionnaires 49
Woman's Questionnaire 49
Man's Questionnaire 52

Appendix B: Example National Historical Events Calendar Used for Date Estimation during the 2012-13 Pakistan DHS **55**

Appendix C: List of Surveys Analyzed in Chapter 2 ..... **57**

Appendix D: List of Surveys Analyzed in Chapter 3 ..... **61**

Tables

Table 2.1 Meaning of DHS completeness codes 7
Table 2.2 Surveys with at least 50% incompleteness in women's reported birthdate/age 11
Table 2.3 Surveys with at least 50% incompleteness in women's reported age at first union 12
Table 2.4 Surveys with at least 50% incompleteness in men's reported birthdate/age 12
Table 2.5 Surveys with at least 10% heaping in reported age from the household survey 15
Table 2.6 Surveys with at least 10% heaping in reported age from the woman's interview 16
Table 2.7 Surveys with at least 10% heaping in reported age from the man's interview 16
Table 2.8 Surveys with age transfers from age 15 to 14 outside the range of -15% to +15%, and the statistical significance of those transfers 20
Table 2.9 Surveys with age transfers from age 49 to 50 outside the range from -15% to 50%, and the statistical significance of those transfers 21
Table 2.10 Surveys with average transfers for health that are 15% or greater and statistical significance of those transfers 22
Table 2.11 Countries in the poorest quintile of data quality for ages and dates, based on a combination of standardized scores for incompleteness, heaping, and displacement in DHS surveys during 2000-2015 25
Table 2.12 Countries in the highest quintile of data quality for ages and dates, based on a combination of standardized scores for incompleteness, heaping, and displacement in DHS surveys during 2000-2015 26

Figures

Figure 1.1 Procedures for the collection of an individual's age and date of birth from The DHS Program's interviewer manual 3
Figure 2.1 Histograms showing the level of any type of incompleteness of reported women's age, women's age at first union, men's age, or children's ages 11
Figure 2.2 Scatterplots showing the incompleteness of women's age at first union, men's age, or children's age, versus the incompleteness of birthdates/ages of women 13
Figure 2.3 Histograms showing the level of heaping in reported age in the household survey, in the survey of women, in the survey of men, and of children in the birth histories 15
Figure 2.4 Scatterplots showing the level of heaping in reported age in the survey of women, the survey of men, and the children in the birth histories, versus the heaping in reported age in the household survey 17
Figure 2.5 Histograms showing the level of displacement from age 15 to 14, or from age 49 to 50, for women in the household survey, and displacement across the boundary for the health questions in the birth histories 19
Figure 2.6 Scatterplot showing the level of displacement from age 15 to 14 versus the level of displacement from age 49 to 50, for women in the household survey 22
Figure 2.7 Changes in the levels of incompleteness, heaping, displacement, and a combination of all three in DHS surveys conducted from 2000 to 2015 24
Figure 2.8 Distribution of the summary measure of incompleteness, heaping, and displacement across the 148 surveys in this analysis 25
Figure 3.1 Distribution of the share of marriages by direction of age difference 28
Figure 3.2 Distribution of average difference between wife's estimate of husband's age and husband's self-reported age, by marriage type 30
Figure 3.3 Distribution of average difference between husband's estimate of wife's age and wife's self-reported age, by marriage type 31
Figure 3.4a Myers' Index for men's ages, self-reports versus spousal estimates 32
Figure 3.4b Myers' Index for men's ages, self-reports versus spousal estimates (continued) 33
Figure 3.5 Myers' Index for women's ages, self-reports versus spousal estimates 34
Figure 3.6 Coefficient for wife's self-reported age from regression on absolute difference between husband's self-reported age and wife's estimate of husband's age 35
Figure 3.7 Coefficient for husband's self-reported age from regression on absolute difference between wife's self-reported age and husband's estimate of wife's age 36
Figure 3.8 Coefficient for wife having primary education (versus none) from regression on absolute difference between husband's self-reported age and wife's estimate of husband's age 37
Figure 3.9 Coefficient for wife having secondary or higher education (versus none) from regression on absolute difference between husband's self-reported age and wife's estimate of husband's age 38
Figure 3.10 Coefficient for husband having primary education (versus none) from regression on absolute difference between wife's self-reported age and husband's estimate of wife's age 39
Figure 3.11 Coefficient for husband having secondary or higher education (versus none) from regression on absolute difference between wife's self-reported age and husband's estimate of wife's age 40

Preface

The Demographic and Health Surveys (DHS) Program is one of the principal sources of international data on fertility, family planning, maternal and child health, nutrition, mortality, environmental health, HIV/AIDS, malaria, and provision of health services.

One of the objectives of The DHS Program is to continually assess and improve the methodology and procedures used to carry out national-level surveys as well as to offer additional tools for analysis. Improvements in methods used will enhance the accuracy and depth of information collected by The DHS Program and relied on by policymakers and program managers in low- and middle-income countries.

While data quality is a main topic of the DHS Methodological Reports series, the reports also examine issues of sampling, questionnaire comparability, survey procedures, and methodological approaches. The topics explored in this series are selected by The DHS Program in consultation with the U.S. Agency for International Development.

It is hoped that the DHS Methodological Reports will be useful to researchers, policymakers, and survey specialists, particularly those engaged in work in low- and middle-income countries, and will be used to enhance the quality and analysis of survey data.

Sunita Kishor
Director, The DHS Program

Abstract

This report assesses the quality and consistency of age and date reports in DHS surveys conducted since 2000 in 67 countries. It is the most recent of several reports on various aspects of DHS data quality.

The first chapter describes the steps of editing and imputing during fieldwork and data processing. Great care is taken to train and supervise the interviewers to obtain the best possible estimates of ages and dates. The eligibility of adults for the surveys of women and men depends on obtaining accurate values of age near the lower and upper age boundaries within the household survey. The eligibility of young children for the detailed health questions depends on obtaining accurate estimates of when they were born within the surveys of women. Age-specific fertility rates, under-five mortality rates, immunization rates, anthropometry scores, and many other DHS indicators depend on accurate estimates of age.

An appendix provides an inventory of all the locations in DHS surveys where the respondents are asked for ages and dates. This assessment focuses on just a few of those locations: the ages of all household members, provided by the household respondent during the household survey; the self-reported ages and birthdates of women and men in the surveys of women and men; women's self-reports of age and date of first union in the survey of women; the birthdates (and ages, if living) of children in the birth histories, provided by the mother; and the women's and men's estimates of their respective spouses' ages in the surveys of women and men.

The second chapter assesses the ages listed above, other than spousal estimates, in terms of three types of measures: incompleteness, heaping, and transfers. A total of 11 indicators are used. For each indicator, the distribution across all surveys is described and the surveys with the most extreme levels are identified. All of these measures vary substantially. There are many surveys with values close to zero on all measures, and others with very high values. There are some surveys in which month of birth is hardly ever given. Age/date transfers are sometimes large but in a direction opposite to what we would expect, particularly around age 15 or around the date for the health questions, clearly as a result of over-correction during training and supervision. Surveys with extreme values are listed. Summary indices of incompleteness, heaping, and transfers are constructed and tracked over time. The indicators fluctuated substantially from 2000 to 2015 and did not show a systematic trend. A single composite index is constructed for each of the 67 countries. The countries in the highest quintile (with the most problems) and the lowest quintile (with the fewest problems) are identified.

The third chapter investigates the quality and consistency of spousal age estimates compared to self-reports in the surveys that included interviews of men and where women and/or men were asked to estimate the age of their spouse(s). In the absence of an age gap between spouses, women tend to estimate that their husbands are older than their self-reported age and men tend to estimate that their wives are younger than their self-reported age. Evidence indicates that where there is an age difference between spouses, women tend to estimate in a way that reduces the gap: they underestimate the age of older husbands and overestimate the age of younger husbands. Men underestimate the age of wives who are older than they are, which reduces the gap, but they also tend to underestimate the age of wives who are younger, which increases the gap. In the vast majority of countries, there was more heaping for estimates of spouse's age than for self-reported age.

There is evidence that displacement and heaping, in particular, can be reduced, through training and supervision, but there is also evidence that too much focus on displacement of children or on heaping at final digit 0 can lead to over-correction. The biggest determinant of good age reporting is probably the value, to the respondents, in everyday life, of knowing their ages or the ages of their children. This component of data quality varies from one setting to another and is outside the control of a survey operation.

1 Background

Surveys conducted by The Demographic and Health Surveys Program (DHS) are a principal source of data for vital rates and health in developing countries. Ages and dates are foundational to a number of key demographic indicators computed by The DHS Program. Periodic assessments of the quality of age and date data are an important element of program monitoring and can help improve data quality. To that end, this report examines the quality and consistency of age and date reporting in DHS surveys conducted since 2000. It includes household, individual, and couples reports of ages and dates as well as birth histories. We begin with an overview of ages and dates collected by DHS surveys, a description of the process of editing and imputation, in the field and during data processing, and prior study findings about age and date quality of DHS data.

1.1 DHS Age and Date Questions Analyzed in this Report

The DHS Program has fielded multistage household surveys in low- and middle-income countries since 1985. In these surveys, after sampling and household selection are complete, interviewers approach each eligible household for an interview with an adult respondent[1] for the household using the household questionnaire, which includes an inventory of all household members. Household members are defined as individuals who are usual residents as well as individuals who stayed with the household as guests the previous night. Based on the household inventory, interviewers determine which women in the household are eligible for individual interviews and—if the survey includes men and the household is selected for male interviews—which men are eligible for individual interviews. If eligible women and men consent to individual interviews, they are conducted separately, typically within a few days of the household interview.

In the period covered by this report—from 2000 until present—The DHS Program has spanned four phases: DHS-4 (1997-2003), DHS-5 (2003-2008), DHS-6 (2008-2013), and DHS-7 (2013-present). The DHS Program publishes core questionnaires for each phase. Household respondents and individual women and men are asked a number of questions related to ages, dates, and durations of time. A full list of these questions and a description of their evolution is included in Appendix A. Below we describe the survey questions used for analysis in this report.

1.1.1 Household Questionnaire

All core household questionnaires from DHS-4 through DHS-7 (ICF International 2011, 2015b; Macro International 2008a; ORC Macro 2001a, 2001b) ask an adult household respondent to provide the name (Q2)[2] and age (Q7) of all household members, to identify a head of household, and to describe each member's relationship to the household head (Q3). Specifically, the household respondent is asked *"How old is (NAME)?"*

1.1.2 Woman's Questionnaire

All woman's core questionnaires from DHS-4 through DHS-7 (ICF International 2011, 2015d; Macro International 2008c; ORC Macro 2001a, 2001b) ask women the following items analyzed in this study:

[1] The respondent for the household questionnaire may be any resident adult who consents to and is capable of answering the questions.

[2] Question numbers given here and in subsequent sections are from DHS-4 Model "A" questionnaire (ORC Macro 2001a). Questions may be numbered differently in subsequent rounds.

- Birthdate: *"In what month and year were you born?"* (Q105)
- Age: *"How old were you at your last birthday?"* (Q106)
- Age of each child who is still alive: *"How old was (NAME) at his/her last birthday?"* (Q217)
- Age at first union—if date of first union is known, this is computed from date of birth and date of first union; if unknown, respondents are asked age at first union:
  - If married or lived with man as if married only once: *"In what month and year did you start living with your husband/partner?"* (Q511)
  - If married or lived with a man as if married more than once: *"Now we will talk about your first husband/partner. In what month and year did you start living with him?"* (Q511)
  - If the year first started living with first husband/partner is unknown: *"How old were you when you first started living with him?"* (Q512)
- If currently married or living with a man as if married, *"How old was your husband/partner on his last birthday?"* (Q702)

1.1.3 Man's Questionnaire

Although DHS surveys included men as early as 1987, the DHS core questionnaires did not include a model man's questionnaire until DHS-5. In the DHS-5 model man's questionnaire and in all subsequent man's questionnaires through DHS-7 (ICF International 2011, 2015c; Macro International 2008b), men are asked the following questions related to age and dates that are used in this report:

- Birthdate: *"In what month and year were you born?"* (Q106)
- Age: *"How old were you at your last birthday?"* (Q107)
- Age of all current wives/partners: *"How old was (NAME) on her last birthday?"* (Q408)

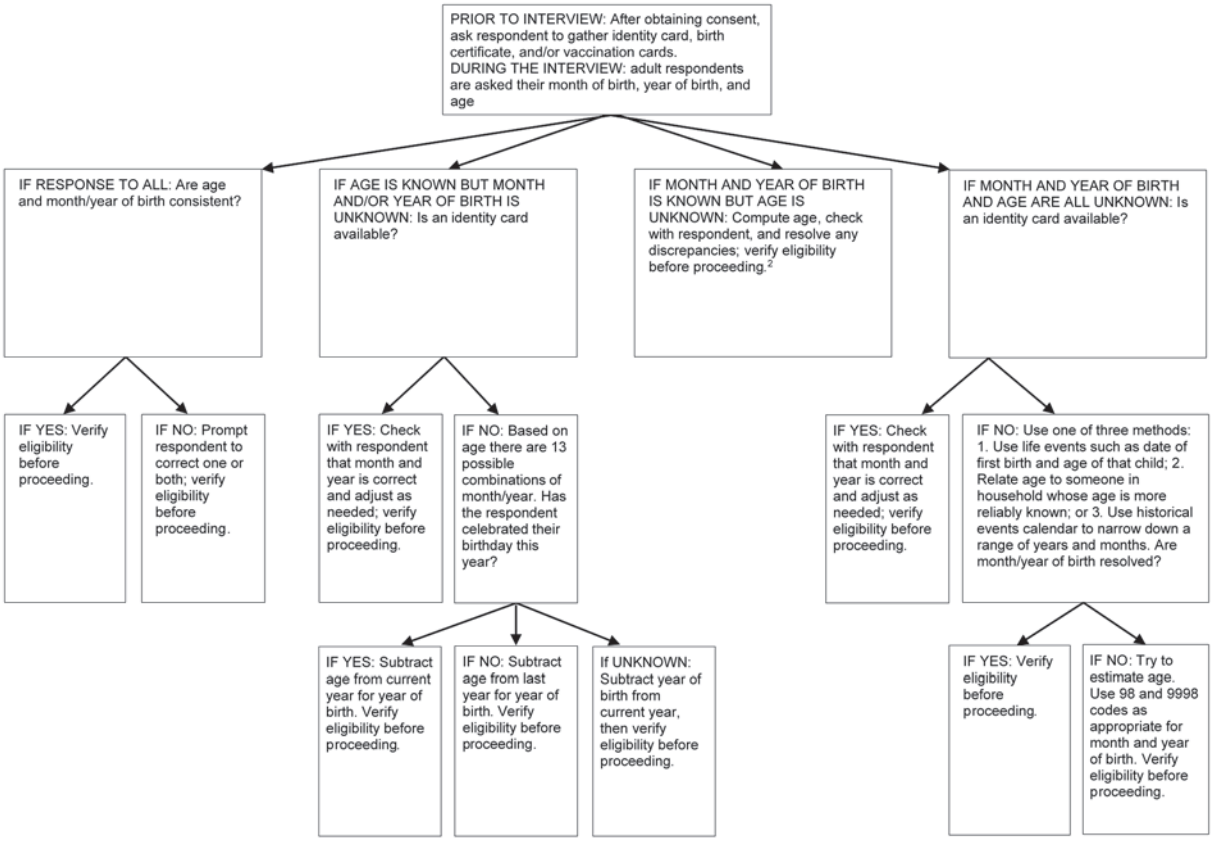
1.2 The DHS Program's Procedures for Obtaining Accurate Age and Date Information

The DHS Program has established a number of field procedures to obtain accurate information on age and date of birth from respondents. Such procedures are of particular importance in countries and areas where innumeracy is prevalent and vital recordkeeping is deficient. Prior to each survey, interviewer training typically lasts several weeks, with modules about age and date reporting that include practical exercises. The DHS Program's interviewer manuals (ICF 2017; ICF International 2012) contain a number of instructions for interviewers on obtaining complete and accurate age and date information from individual respondents. Figure 1.1 illustrates these procedures. Interviewers are also instructed to use these methods for the household age roster if the age of any household member is unknown. In countries lacking vital registration systems, identity cards may themselves be inaccurate. For this reason, as noted in Figure 1.1, if interviewers ask for a card in the process of establishing date of birth they are advised to check with the respondent about the veracity of the date printed on the card. When both age and date of birth are unknown, interviewers are instructed to reference respondent's age at particular life events (for example, birth of first child) and number of years ago the event took place, to relate their age to someone in household whose age is more reliably known, or to use the local and/or national historical events calendar they received during training. An example of a national historical events calendar distributed to interviewers for use in date of birth estimation is given in Appendix B.

In practice, during individual interviews the process of obtaining ages and dates may be slightly different from what is shown here; in particular it may be more iterative. While interviewers are explicitly instructed

to ask age and date of birth independently from the household interview, if there is uncertainty about the respondent's age then the age assignment from the household interview may be taken as a starting point for a discussion about the individual respondent's age and date of birth.

Figure 1.1 Procedures for the collection of an individual's age and date of birth from The DHS Program's interviewer manual



1. This figure is a summary only; see ICF 2017 and ICF International 2012 for more detailed instructions and procedures.
2. Check that the respondent meets the age range required to qualify for an individual interview.

Oversight and quality-checking are important aspects of ensuring the accuracy of age and date reporting. In the absence of adequate supervision there may be a tendency toward age displacement that reduces the time burden on interviewers and respondents. For example, if a household member is outside the range for individual interviews (typically age 15-49), an individual interview is not administered; if a child is older than age 5, the interviewer can skip a battery of questions about the child's circumstances of birth, health, and nutrition. DHS field supervisors and survey managers are cognizant of this motivation for a possible bias. Field editors and supervisors may ask interviewers to return to a household where ages and dates of birth appear overly rounded or are slightly outside of the thresholds. Supervisors routinely examine field check tables for age heaping and may return to a household or to an entire survey cluster in an attempt to obtain more precise information.

1.3 The DHS Program's Procedure for Editing Inconsistent and Imputing Incomplete Dates

After data collection is complete, The DHS Program follows a standard procedure for checking the consistency of dates and for imputing missing dates (and, by extension, ages). These procedures are described in detail in the DHS data editing and imputation manual (Croft 1991). Inconsistent dates are flagged and incomplete dates are imputed for the following demographic events:

- Date of birth of the respondent
- Date of first union
- Date of birth of each child
- Date of conception and current pregnancy
- Date of sterilization of respondent or partner

Data editing and imputation require the creation of an event table involving the aforementioned events plus employing ancillary data such as the duration of breastfeeding, amenorrhea, and abstinence after the birth of a child. These pieces of information create logical constraints on demographic events; for example, a minimum seven-month interval between non-multiple births, the impossibility of becoming pregnant during a period of postpartum abstinence, and so forth. Inconsistent data are flagged and resolved prior to imputation. Imputation is a four-stage process during which logical ranges for the date of each event are established and iteratively narrowed based on the sequence of events until a final random value is assigned. Imputed dates are flagged in final DHS datasets.

1.4 Prior DHS Studies on Data Quality

Given the essential need for correct and consistent age and date data for the computation of demographic indicators, The DHS Program and its predecessor, the World Fertility Survey, have undertaken a number of studies on age and date data quality. In an early study, Arnold (1991) found that eligibility errors—including birth displacements past the five-year line—may affect fertility rates and child and infant mortality rates by a few percentage points. However, the overall quality of DHS-I data was quite good. Rutstein et al. (1990) found similar issues with misplaced eligibility and birth displacement and similar probable effects on demographic rates. Curtis (1995) examined the quality of data used to compute infant and child mortality in DHS surveys and found that the date of birth was consistently less complete for dead children than for surviving children but did not substantially affect mortality rates. Gage (1995) examined the quality of DHS-II data on the age at first union, at first birth, and at first sexual intercourse and found evidence of heaping in age at first union and years of first birth, but not a directional bias.

More recently, a DHS report on age and date quality (Pullum 2006) examined nine symptoms of poor data quality in DHS surveys from 1985 to 2003. There was evidence of age heaping and age displacement, but an overall improvement in data quality over the time period studied. A second study that examined fieldwork found that data quality was worse in rural areas, at the beginning and end of fieldwork, and in instances where a translator was used (Johnson et al. 2009). Schoumaker (2014) examined the quality and consistency of DHS fertility estimates and found that most estimates were of good or acceptable quality but that not all should be taken at face value due to the omission of recent births, a tendency to report the first birth as more recent than it was (the Potter effect), and differences in sample composition. Pullum and Becker (2014) examined birth histories to study the effects of the omission of births and/or deaths, potential displacement, and misreporting. They found that the quality of recent surveys was high; omission and displacement of births is typically around 2%, while omission and displacement of deaths may be up to 5%.

1.5 Report Structure

This report examines the quality of age and date reporting in DHS surveys conducted during the interval from 2000 to 2015[3]. Chapter 2 reviews 148 surveys conducted in 67 countries, using four indicators of

[3] The fieldwork for two surveys, Angola 2015-16 and Myanmar 2015-16, extended into 2016, but it would be misleading to include 2016 in the reference interval of time. For all surveys included in this report, the range of years for the start of fieldwork was 2000 to 2015.

incompleteness, four indicators of heaping, and three indicators of displacement. All of the indicators are expressed as percentages that ideally would be at a level of 0%, except for possible random variation. Incompleteness refers to ages that should be accompanied by the month and year of birth (except for women's age at first union, which should be accompanied by the month and year of first union). If any component of age, month, and year is missing or the combination provided is inconsistent, then the response is considered incomplete. Heaping refers to digit preference, most commonly for final digits 0 or 5 in reported age. Displacement refers to an irregularity in the age distribution that suggests the interviewer has tended to shift women out of the eligible age range 15-49 for the survey of women, or children out of the eligible range of birthdates for the questions on child health.

Chapter 2 also describes the distribution of the level of each indicator across countries and identifies the specific surveys with the highest levels. It also combines the indicators into groups to summarize the levels of incompleteness, heaping, and displacement for each survey, and it describes year-to-year levels from 2000 to 2015. Finally, a single composite indicator is used to identify the countries in the worst quintile or in the best quintile of age and date reporting, so that problems in future surveys can be better anticipated.

For each DHS survey that includes a survey of men as well as a survey of women, a "couples' file" is constructed, consisting of a single record for each co-resident woman and man who self-identify as partners. For women—and in some cases, men—the data include the person's own age and the person's report (estimate) of their partner's age. Chapter 3 considers another aspect of the quality of age reports—the correspondence between a respondent's reported age and their spouse's estimate of the respondent's age in 113 surveys for men and 67 surveys for women. The correspondences and differences between the two statements of age for each partner are examined systematically, with reference to a set of hypotheses. Heaping of the spousal age estimate is compared with heaping of the partner's self-reported age. Characteristic patterns of discrepancies are identified. In this chapter, two covariates—age and education—are included.

The report also includes, as Appendix A, a comprehensive review of all ages, dates, and time durations asked in the core questionnaires from DHS phases 4 through 7, and their evolution over time.

2 Incompleteness, Heaping, and Displacement in DHS Surveys

2.1 Data and Methods

This chapter presents a systematic review of misreporting of ages and dates using virtually all of the DHS surveys conducted between 2000 and the closing date for this report—148 surveys in 67 countries. These numbers do not include Malaria Indicator Surveys (MIS) or AIDS Indicator Surveys (AIS) or so-called Interim Surveys but do include groupings of rounds of the Continuous Surveys in Peru and Senegal. The total number of cases is approximately 7.98 million persons of all ages in the household surveys, 1.67 million women in the surveys of women, 0.72 million men in the surveys of men, and 3.79 million children in the birth histories.

Misreporting will be described with indicators of incompleteness, heaping, and displacement. For each of these, three or four indicators will be employed, a total of eleven, variously using the household survey, the survey of women, the survey of men, and the birth histories. We will propose a synthesis of these indicators that can serve as a general indicator of misreporting, and will describe overall trends since 2000.

Incompleteness applies to an instance in which the respondent provides less information about an age or date than is expected, requiring some imputation during data processing, or when there is an inconsistency between the types of information provided that requires some reconciliation. There are eight possible codes, representing the possible combinations of age, year of birth, and month of birth, being provided (and consistent) or not provided (or not consistent).[4] Definitions of incompleteness codes below are from the DHS-6 recode manual (MEASURE DHS and ICF International 2013):

Table 2.1 Meaning of DHS completeness codes

DHS completeness code	Meaning
1: month and year - information complete	Both month and year of the event were specified so no imputation was necessary.
2: month and age - year imputed	The year of the event was not given, but the month of the event and the age of the respondent or child or, in the case of the date of first union, the respondent's age at first union were specified. In most cases this information uniquely identifies the exact date of the event. In a few cases the year of the event was imputed from a choice of two possible years.
3: year and age - month imputed	The year of the event, but not the month, and the age of the respondent or child or, in the case of the date of first union, the respondent's age at first union were specified and only the month of the event was imputed.
4: year and age - year ignored	The year of birth, but not the month, and the age of the respondent or child were specified. In surveys where it is believed that the year of birth was calculated by just subtracting age from the current calendar year, year and may be imputed within a fully twelve-month range that is consistent with stated years of age and date of interview.
5: year - age/month imputed	The year of the event was given but the month of the event was not specified, and neither was the age. The month of the event was imputed.
6: age - year/month imputed	Neither the month nor the year of the event was specified, but age was given and the year and month of the event were imputed from the age.
7: month - age/year imputed	Only the month of the event was given, without the year or age. The year of the event was imputed from other information.
8: none - all imputed	No information was given concerning the date of the event. But month and year of the event were imputed from other information.

[4] In the birth histories for the most recent surveys, day of birth is also asked. If the respondent can provide day, month, and year of the child's birth, then the relevant completeness variable (b10) is assigned code 0. In this report, the only affected surveys are Malawi 2015-16 and Tanzania 2015-16.

Type 1 is most desirable. Several of the other logically possible combinations hardly ever occur. Types 3, 6, and 4, in that order, account for virtually all of the incomplete cases. The general rule during imputation is to give higher priority to stated age than to stated year or month of birth. If age in years is given, then there are typically 12 possible months of birth, spread across two adjacent calendar years of birth, that would be consistent with the month and year of interview. Imputation is typically at random within this range. In this analysis, we simply aggregate types 2-8 as a percentage of the total for types 1-8 and call this the "percentage incomplete." It will be calculated for: age in the women's survey (v014 is the incompleteness code for age, v012); age at first union in the women's survey (v510 is the incompleteness code for age at first union, v509); age in the men's survey (mv014 is the incompleteness code for age, mv012); and age of a child in the birth history (b10 is the incompleteness code for month and year of birth and current age if alive). Each completeness code refers to a combination of age, month, and year.

Heaping refers to a tendency to provide an age that disproportionately has specific final digits, typically 0 or 5, although other final digits such as 2 or 8 can also occur more often than would be expected. The easiest way to describe heaping or digit preference is to calculate the percentage of cases with each possible final digit (0, 1, …, 9), add up the absolute deviations of those percentages from 10%, and divide the total by two. This indicator is an application of the Index of Dissimilarity, which summarizes the deviation of any observed categorical distribution from a distribution that is "expected" under some model. The index is interpreted as the percentage of cases that would have to be shifted from an over-represented category to an under-represented category in order to match the "expected" distribution. The division by two is required because when one case is moved it will simultaneously reduce the excess by one and reduce the deficit by one.

A longstanding indicator of age heaping is Myers' Blended Index (herein Myers' Index). This index is a minor modification of the Index of Dissimilarity that takes into account the overall pattern of most observed age distributions, such that the percentage of the cases at age x+1 tends to be less than the percentage at age x. Such a pattern results from the cumulative impact of mortality, as well as, often, an increasing number of births every year compared with the previous year. For that reason, even in the absence of digit preference, there would tend to be fewer cases with final digit 1 than with final digit 0, fewer cases with final digit 2 than with final digit 1, etc., and fewer cases with final digit 9 than with final digit 8. Myers' Index adjusts the Index of Dissimilarity for that gradient. For either index to be applied correctly, the age range must be a multiple of 10 years. We will calculate this index for age in the household survey (hv105, using ages 0-79), age in the survey of women (v012, using ages 20-49), age in the survey of men (mv012, using ages 20-49), and age of living children in the birth histories (b8, for ages 0-29).

It would be possible to examine digit preference in the stated calendar year of birth, rather than age at the time of the survey. For many surveys, for example, there appears to be a preference for calendar years ending in 0, particularly for the year 2000. This will not be investigated systematically here, partly because heaping on calendar years would be confounded with heaping on age for interviews conducted in calendar years ending with 0 or 5, that is, in 2000, 2005, 2010, and 2015.

Incompleteness and heaping probably tend to result from the respondent genuinely not knowing the age in question. The procedure for field imputation of ages and dates was described in Chapter 1. In many surveys, some assistance is provided in the form of identity cards, reference events, and relative ages. But if the respondent really does not know the birth year or age, there is a limit to what can be done during the interview to obtain a correct response.

Displacement—the third potential type of age misreporting considered here—refers to a shifting of age or birthdate across some relevant boundary. Three examples of potential displacement will be considered. The first is a possible transfer from age 15 to age 14 for females in the household survey. One function of the household survey is to ascertain eligibility for the survey of women. The lowest age of eligibility is 15 (with

very few exceptions). If the age of a woman is reported to be 14, rather than 15, then she will not be eligible for the household survey. The second example is a possible transfer from age 49 to age 50 for women in the household survey, which will also remove women from eligibility because the upper age limit is 49. Note that transfers from age 49 to 50 related to the eligibility criterion are confounded with heaping at age 50. Heaping at age 50, as such, would tend to be drawn from true ages just above or just below 50. Transfers out of eligibility for the survey of women would only displace from true ages below 50. Nevertheless, transfers from age 49 to 50 can arise from either source.

Another example of displacement is a possible transfer of children in the birth histories across the boundary for eligibility for the child health questions. Eligibility is normally determined by whether the child was born within the five years before the month of interview, or, more precisely, in January of the fifth calendar year before the beginning of fieldwork, or later.[5] The first calendar month for the health questions is the same for all interviews and is specified in the data files (with a century month code) as v017. Some health data are dropped during data processing if they refer to the interval between that specific month and the 60th month before the month of interview. For a few surveys, the interval for the health questions is less than five years; for a few surveys, v017 is a month other than January. These potential transfers are estimated with a data file that lists all births in the birth histories (the BR file), including those that are reported with a birthdate before month v017, as well as those during and after month v017, which were eligible for the health questions.

The three age/date transfers would be attributable to the interviewer, motivated by a possible desire to reduce the required number of in-depth interviews. At the point when the interviewer asks these questions about age, the respondent does not know that the responses will be used to determine eligibility for further questioning. The interviewer, of course, does know how they will be used. Nevertheless, on the basis of training and field observations, it is generally believed that the tendency of interviewers to displace an age/date in a direction that would avoid extended interviews occurs mainly when there is ambiguity or uncertainty from the respondent.

The indicator of potential age/date transfers was first used in Methodological Report #5 (Pullum 2006, pp. 81-82). It is based on an assumption that the true number of cases in a short sequence of four successive years—two of which are before the boundary and two of which are after the boundary—is linear on a log scale. Another way to describe the expected regularity is that the ratios of the frequencies at successive years should all be equal. Referring to the four successive years of age generically, as years 1, 2, 3, and 4, the assumption is that the true number of cases at year 2, divided by the number at year 1, is the same as the number at year 3 divided by the number at year 2, or the number at year 4 divided by the number at year 3. We also assume that the displacement only involves years 2 and 3 and that the total for years 2 and 3 is correct.[6] These assumptions provide leverage to estimate the percentage of true cases in year 3 that were shifted downward to year 2, or conversely, the percentage of true cases in year 2 that were shifted upward to year 3.[7]

When this method is applied to ages 13, 14, 15, and 16, the expected direction of transfers is downward, and a positive sign for the shift will indicate evidence of a transfer in the expected direction. When applied to ages 48, 49, 50, and 51, the expected direction of transfers is upward, and a positive sign for the shift will indicate evidence of a transfer in the expected direction. When applied to child transfers, year 1 is the second year before v017, year 2 and year 3 straddle the boundary, and year 4 is the second year after v017;

[5] This potential transfer is sometimes described as being across the fifth birthday, that is, from age 4 to age 5. The transfer is actually across a specific calendar month and calendar year of birth, as described.
[6] Displacement, like rounding, is probably over a wider range than adjacent years. This is a simplifying assumption.
[7] A detailed description will not be given here; a link is provided in the References.

the expected direction of transfers is downward (from year 3 to year 2), and a positive sign for the shift will indicate evidence of a transfer in the expected direction.

Heaping and displacement are closely related. For both of them, the methods are based on comparing the observed frequencies with the frequencies that would be expected under a model that smooths the data.

The data files include other distributions of ages and dates that could be investigated for heaping or displacement. For example, another outcome that has been included in previous assessments of data quality is heaping of age at death on exactly 12 months. Age at death is asked in months for children who died before two years of age, and asked in years for age 2 and above. There is usually a tendency for some responses to occur disproportionately at exactly 12 months, probably through a thought process that converts a less precise "one year" to more precise "12 months". In terms of completed years and months, however, the range of 0-11 months is understood by DHS to refer to age 0 years (infancy), and 12 months is the first (completed) month at age 1. As an illustration, in the Afghanistan 2015 survey, among children born in the past five years the numbers of deaths reported at 10, 11, 12, 13, and 14 completed months of age is 11, 31, 76, 14, and 6. It is clear that many of the 76 deaths reported at exactly 12 months should have been placed at earlier months or later months. The displacement of deaths from age 0 to age 1 resulted in a slight underestimate of recent infant mortality (and over-estimate of mortality for age 1 to age 4). There are a number of possible ways to measure this type of reporting error, interpreting it as a result of heaping or displacement, but the weakness of any measure based on this particular type of error is that it involves small frequencies and more sampling error than the other indicators. For that reason, it is not included in this report.

For each measure of incompleteness, heaping, and displacement, we will describe the distribution across all surveys since 2000 and will identify the surveys with the most extreme values of the measures. All measures treat all observations equally, ignoring sampling weights.

2.2 Incompleteness

Incompleteness of age and date reporting will be assessed for two ages in the survey of women—the woman's stated age and age at first union; for one age in the survey of men—the man's stated age; and for one age in the birth histories—the woman's report of the child's birth date and age. The household survey is not included for this type of indicator.[8] As stated earlier, the indicator is the percentage of cases for which the age, year of birth, and month of birth are not all provided or are not all consistent with the month and year of interview. Figure 2.1 gives the distribution of this percentage for the four indicators, across all surveys. The four graphs in the figure have the same horizontal scales, the percentage of responses that are incomplete, but different vertical scales.

When age is incomplete, it is usually because age and year of birth were provided, but not a month, or the month was inconsistent with the year and age, so the month had to be imputed or modified. It is rare for age itself to require imputation. Over all the surveys, the average level of incompleteness is 21.8% for women's birthdate/age. Incompleteness of woman's age in excess of 50% was found in the 22 surveys listed in Table 2.2. Overwhelmingly, the incompleteness was only in terms of month, but a threshold of 50% is clearly a high level of any kind of incompleteness.

[8] The age variable in the household survey, hv105, is not accompanied by a month and year of birth, or by a separate indicator of completeness, and is not subject to imputation or editing during data processing. The variable has a potential code for "missing," 98, but this code is rarely required and has little diagnostic value. The greatest use of this code, by far, was in the Angola 2015-16 survey, in which it was assigned to 1.2% of the respondents. It was assigned to 0.3% of the cases in the Mozambique 2000 survey and 0.2% of the cases in the Namibia 2000 and Senegal 2010-11 surveys. In all other household surveys the level is 0.1% or less.

Figure 2.1 Histograms showing the level of any type of incompleteness of reported women's age, women's age at first union, men's age, or children's ages

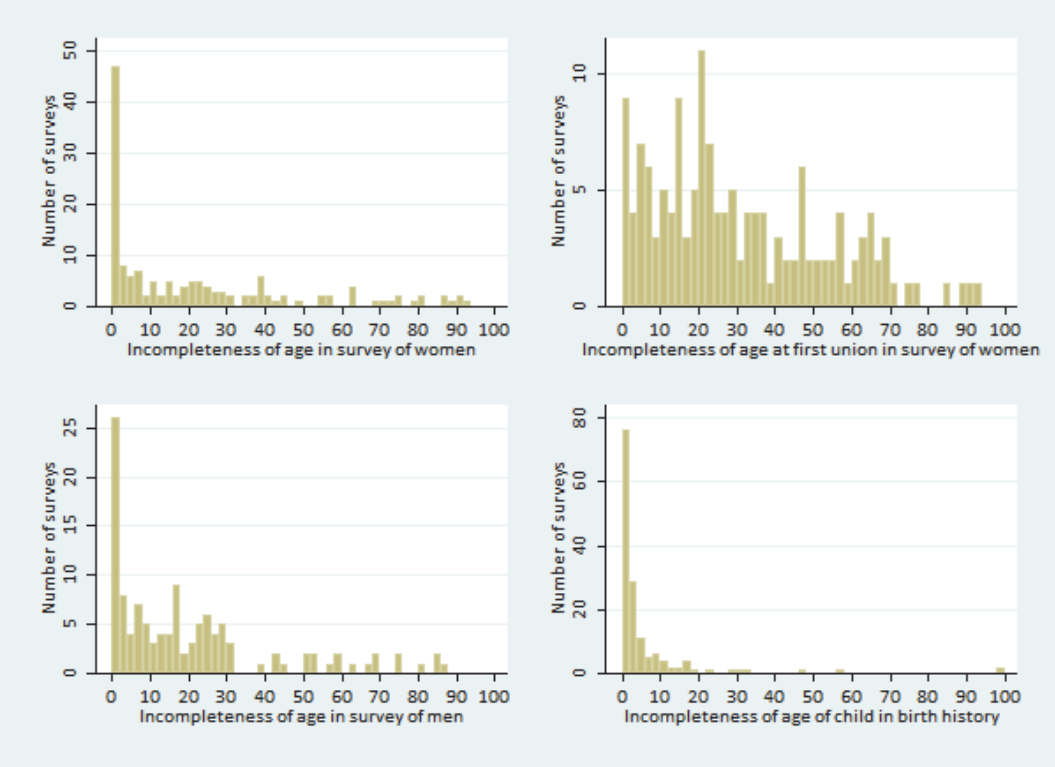


Table 2.2 Surveys with at least 50% incompleteness in women's reported birthdate/age

Survey	%
Rwanda 2005	55.0
Burundi 2010	55.1
Nepal 2001	56.2
Ethiopia 2005	57.2
Niger 2006	62.1
Ethiopia 2011	63.3
Chad 2014-15	63.5
Benin 2006	63.8
Mali 2006	68.5
Yemen 2013	70.5
Pakistan 2006-07	73.6

Survey	%
Benin 2001	74.1
Pakistan 2012-13	75.6
Ethiopia 2000	79.4
Mali 2001	81.1
Burkina Faso 2003	81.2
Niger 2012	86.3
Bangladesh 2014	86.9
Guinea 2005	89.4
Bangladesh 2011	91.4
Bangladesh 2007	91.8
Bangladesh 2004	93.7

Four of the five surveys with greatest incompleteness of women's age were conducted in Bangladesh (2004, 2007, 2011, and 2014). Each of these four surveys had slightly less incompleteness than the preceding one, suggesting some improvement, but not a great deal. The underlying problem is that many women in Bangladesh simply do not know their birthdates. Several other countries appear more than once, e.g., Niger (2006 and 2012), Mali (2001 and 2006), Ethiopia (2000, 2005, and 2011), and Benin (2001 and 2006).

The average level of incompleteness for age at first union is considerably higher, at 30.8%. Table 2.3 lists the 32 surveys with a level above 50%. In addition to many of the surveys that were on the previous list,

this list includes five surveys from Senegal (2005, 2010-11, 2012-13, 2014, and 2015)—all rounds of the Senegal Continuous Survey since 2010, plus an earlier standard DHS from 2005.

Table 2.3 Surveys with at least 50% incompleteness in women's reported age at first union

Survey	%
Cameroon 2004	51.4
Gabon 2012	51.8
Senegal 2012-13	52.8
Senegal 2010-11	53.6
Ghana 2008	54.3
Madagascar 2003-04	55.9
Senegal 2015	56.0
Cote d'Ivoire 2011-12	56.3
Ethiopia 2011	57.0
Senegal 2014	57.6
Comoros 2012	58.6
Niger 2006	61.3
Bangladesh 2007	61.7
Yemen 2013	62.3
Sierra Leone 2013	62.7
Ghana 2003	63.3

Survey	%
Senegal 2005	64.6
Sierra Leone 2008	65.1
Bangladesh 2014	65.2
Benin 2011-12	65.6
Bangladesh 2011	66.9
Sao Tome and Principe 2008-09	67.9
Ethiopia 2000	68.7
Mali 2006	68.7
Nepal 2001	69.9
Mali 2001	71.9
Chad 2014-15	75.4
Benin 2006	76.3
Benin 2001	84.7
Burkina Faso 2003	88.2
Niger 2012	90.8
Guinea 2005	93.1

For men's birthdate/age, the mean level of incompleteness is 20.8%. Table 2.4 lists the 17 surveys that have a level above 50%. The surveys on this list match closely with the surveys on the list for incompleteness of women's birthdate/age. Some of the differences are simply due to the fact that not every DHS survey includes a survey of men. For example, the Bangladesh 2014 survey did not include a survey of men. The Bangladesh 2004, 2007, and 2011 surveys are at the extreme end of the list, with levels of incompleteness that show a slight increase, rather than decline, from one to the next. The Niger 2006 survey had a level of 50.7%, and the 2012 survey had a much higher level, 80.3%. Three surveys from Ethiopia are on the list. The 2000 survey had a level of 69.4%; the 2005 and 2011 surveys were considerably lower, at 53.2% and 56.8%, respectively. When following successive surveys in the same country, it is difficult to generalize about whether there has been a trend toward less incompleteness or toward more.

Table 2.4 Surveys with at least 50% incompleteness in men's reported birthdate/age

Survey	%
Niger 2006	50.7
Chad 2014-15	51.5
Burundi 2010	52.5
Ethiopia 2005	53.2
Ethiopia 2011	56.8
Benin 2006	58.6
Rwanda 2005	59.3
Benin 2001	63.0
Mali 2006	66.3

Survey	%
Guinea 2005	68.9
Ethiopia 2000	69.4
Burkina Faso 2003	74.7
Mali 2001	75.2
Niger 2012	80.3
Bangladesh 2004	84.3
Bangladesh 2007	84.9
Bangladesh 2011	86.5

The level of incompleteness of birthdates and ages in the birth histories is far less, only 4.6% on average. The only survey above 50% is Guinea 2005.

Across surveys, there is close correspondence between the measures of incompleteness. The correlation between the incompleteness levels for women's age and men's age is 0.97. The correlation between

incompleteness of age at first union and incompleteness of women's age is 0.72. The correlation between incompleteness of age at first union and incompleteness of men's age is 0.67. Whether incompleteness arises from the respondents' own knowledge of ages and dates, or from the quality of the data collection, we would expect high levels of correspondence, because the men and women are in the same households and the interviewers all have the same training and supervision. Incompleteness of children's birthdates and ages has much lower correlation with the other three measures—0.18 with incompleteness of women's age or men's age and 0.21 with incompleteness of age at first union.

Figure 2.2 Scatterplots showing the incompleteness of women's age at first union, men's age, or children's age, versus the incompleteness of birthdates/ages of women

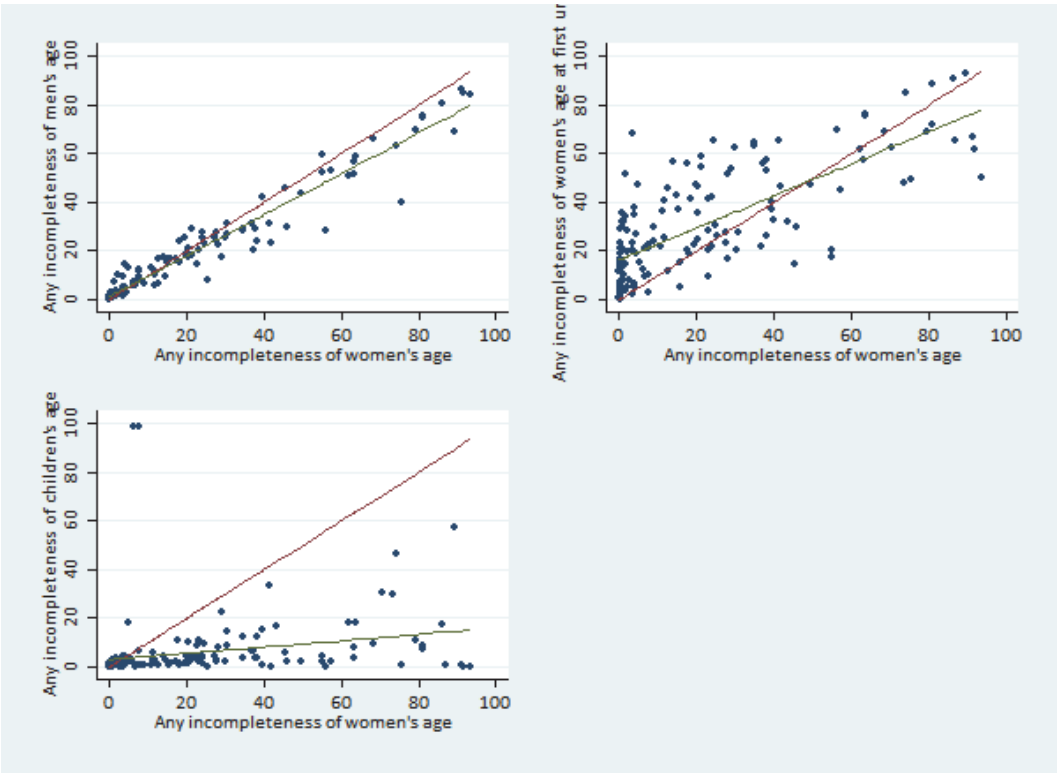


Figure 2.2 provides scatterplots of pairings of these four indicators. The reference indicator is the level of incompleteness of women's age, which is the horizontal axis for all three graphs. In each graph, a red diagonal line represents hypothetical equality of the indicator on the vertical axis with incompleteness of women's age, and a green line is the fitted line when the indicator on the vertical axis is regressed on incompleteness of women's age.

In the upper left graph, the nearness of the two lines is due to the high correlation between incompleteness of men's age and incompleteness of women's age. Far more points are below the line of equality than are above it. That is, given the woman's level of incompleteness, the man's level tends to be lower. A handful of surveys are 20 to 30 points below the line, implying substantially less incompleteness for men than for women. Our interpretation would be that knowledge of birthdate and age may be more salient for men than for women, in the countries represented by the points farthest below the red diagonal.

In the upper right graph, there is much more scatter, and far more points are above the line of equality, rather than below it. That is, incompleteness of women's age at first union is not just more common than incompleteness of women's age—the former is more likely for any given level of the latter. Even so, another handful of surveys are 20 to 30 points below the line, a deviation from the typical pattern. Incompleteness

of women's age at first union is probably due partly to potential ambiguity about the criterion for the first union and partly due to the fact that it is not as useful or salient as the woman's own birthdate and age.

The lower left graph shows that the correspondence between incompleteness of the woman's birthdate/age and her children's birthdates/ages is much less than the other correspondences, and also shows that the level of incompleteness is much lower for children than for adults. This graph includes two points that appear to have extreme levels of incompleteness for children. These points are for the Malawi 2015-16 and Tanzania 2015-16 surveys. These two surveys included day of birth for children (the completeness code is b10=0 if day, as well as month and year, is provided). When that is taken into account, only 4.5% and 1.6%, respectively, of children's birthdates/ages did not include day of birth; only 2.8% and 0.4%, respectively, had incompleteness codes 2-8. That is, taking into account the revised coding, those two surveys had very low incompleteness. In the future, all DHS surveys will obtain day of birth, as well as month and year, for all children in the birth histories, although not for adults.

2.3 Heaping

As described earlier, age heaping is measured with Myers' Blended Index, which can be interpreted as the percentage of cases that would have to be shifted from over-represented final digits to under-represented final digits, with an adjustment to take account of the general gradient in the age distribution. A heaping index of 0 implies perfect uniformity across final digits 0 through 9, but simply because of randomness we would never expect to achieve that lower limit.

The four graphs in Figure 2.3 show the distributions of the heaping index for the four ages used in this analysis: age in the household survey; age in the women's survey; age in the men's survey; and age in the birth histories. The woman's estimate of her husband's age and the man's estimate of his wife's age will be examined separately in Chapter 3. All four graphs have the same horizontal scale, so the differences in distributions appear clearly, but different vertical scales.

Across surveys, the mean level of heaping is 5.3% in the household survey. The mean, median, and maximum of the level of age heaping are substantially lower for age in the household survey than for age in the surveys of men and women. A lower level of heaping in the household survey, for ages 0-79, than in the surveys of men or women, for ages 20-49, would not be expected, but it results from the difference in age ranges. The wider age range in the household survey includes ages below 20, for which there is generally not much heaping. It includes ages 50-79, for which considerable heaping at 50, 60, and 70 is often found, but the number of people age 0-19 is greater than the number age 50+, so the net effect is less heaping within ages 0-79 than within ages 20-49. Age in the surveys of men and women is sometimes a revision of age in the household survey, with additional probing about month and year of birth, but those revisions are usually small.

Twenty-five surveys, listed in Table 2.5, have a level of heaping above an arbitrary threshold of 10%. The highest level, 15.2%, is in the India 2005-06 survey, and other surveys from the Indian Subcontinent are also on this list—two from Pakistan (2006-07 and 2012-13) and three from Bangladesh (2004, 2011, and 2014). There are also three surveys from Nigeria (2003, 2008, and 2013), three from Benin (2001, 2006, and 2011-12), two from Niger (2006 and 2012), two from Sierra Leone (2008 and 2013), and a mix of other countries that appear on other lists.

Figure 2.3 Histograms showing the level of heaping in reported age in the household survey, in the survey of women, in the survey of men, and of children in the birth histories

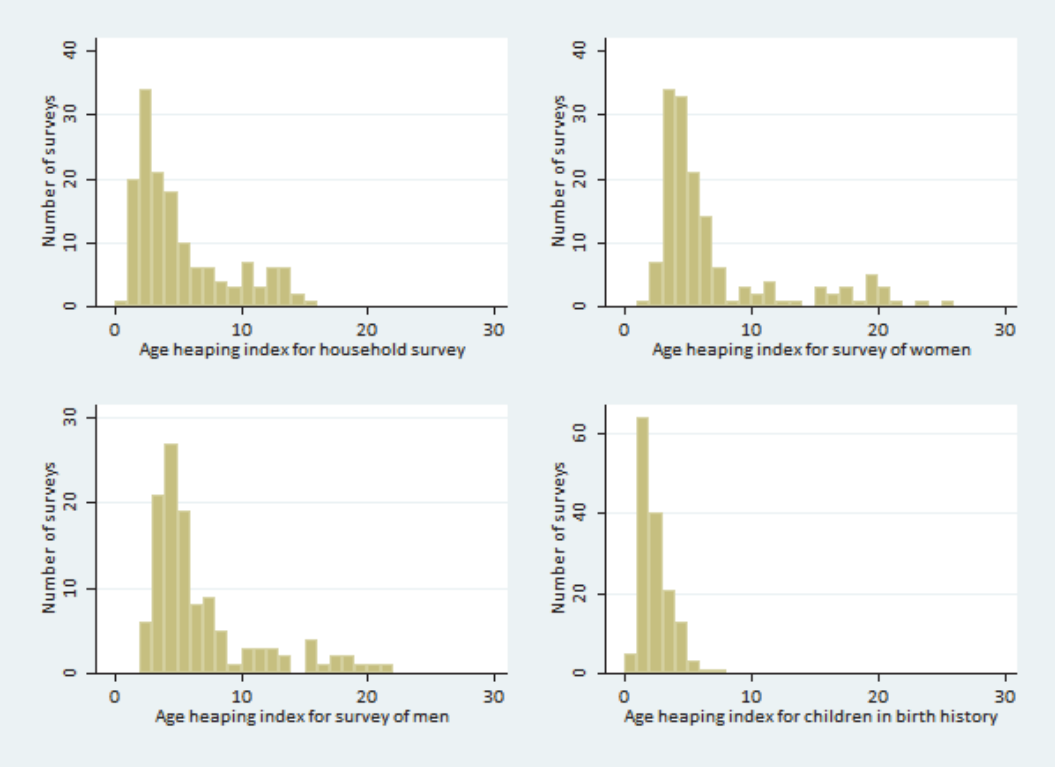


Table 2.5 Surveys with at least 10% heaping in reported age from the household survey

Survey	%
Mali 2012-13	10.3
Yemen 2013	10.5
Bangladesh 2011	10.5
Bangladesh 2014	10.6
Niger 2012	10.9
Bangladesh 2004	10.9
Benin 2001	11.0
Guinea 2012	11.0
Niger 2006	11.2
Guinea 2005	11.5
Chad 2004	12.0
Benin 2011-12	12.1
Sierra Leone 2008	12.4

Survey	%
Benin 2006	12.5
Afghanistan 2015	12.6
Pakistan 2012-13	12.8
Pakistan 2006-07	13.1
Sierra Leone 2013	13.2
Ethiopia 2005	13.2
Nigeria 2013	13.5
Ethiopia 2011	13.7
Chad 2014-15	13.8
Nigeria 2008	14.2
Nigeria 2003	14.3
India 2005-06	15.2

The mean level of heaping is 7.1% in the women's survey. The level is 10% or more in 26 countries, listed in Table 2.6. This list does not include India 2005-06. That survey had a high level of heaping for age in the household survey (15.2%), but only 6.3% for women's age, 7.9% for men's age, and 2.8% for children's birthdates and age. This is an exceptional combination of high heaping on age in the household survey but low heaping in the women's and men's surveys, suggesting that additional attention was given to obtaining better reports of age in the interviews with women and men. The three surveys in Bangladesh that were included on the list for high age heaping in the household survey have also moved off the corresponding

list for the women's survey. The two surveys of Pakistan remain on the list, as do most of the other surveys seen above.

Table 2.6 Surveys with at least 10% heaping in reported age from the woman's interview

Survey	%	Survey	%
Ghana 2008	10.1	Mali 2012-13	17.6
Comoros 2012	10.5	Chad 2004	17.7
Pakistan 2012-13	11.3	Guinea 2012	18.2
Senegal 2010-11	11.4	Benin 2006	19.2
Togo 2013-14	11.4	Niger 2012	19.3
Pakistan 2006-07	11.9	Sierra Leone 2013	19.4
Mali 2001	12.8	Niger 2006	19.4
Ethiopia 2000	13.0	Ethiopia 2011	20.0
Mali 2006	15.4	Guinea 2005	20.2
Benin 2001	15.5	Afghanistan 2015	20.3
Yemen 2013	15.6	Nigeria 2008	20.6
Benin 2011-12	16.6	Ethiopia 2005	21.9
Nigeria 2003	16.6	Sierra Leone 2008	23.2
Nigeria 2013	17.0	Chad 2014-15	25.1

The mean level of heaping in the men's survey is 7.0%. Twenty-three surveys are at or above the arbitrary threshold of 10% and are listed in Table 2.7. Sierra Leone 2008 has the highest level, at 21.1%, and Sierra Leone 2013 is fourth highest, at 18.1%. As with the other lists, Nigeria, Afghanistan, Chad, Guinea, Benin, Ethiopia, and Nigeria are among the countries with the highest levels of heaping. Two surveys from Bangladesh that were below 10% for heaping in the women's survey are above 10% in the men's survey.

Table 2.7 Surveys with at least 10% heaping in reported age from the man's interview

Survey	%	Survey	%
Mali 2006	10.1	Benin 2011-12	15.4
Ethiopia 2000	10.3	Nigeria 2003	15.6
Bangladesh 2007	10.7	Guinea 2005	15.7
Benin 2001	11.1	Benin 2006	16.6
Pakistan 2012-13	11.5	Nigeria 2013	17.2
Mali 2012-13	12.0	Guinea 2012	17.7
Ethiopia 2011	12.6	Chad 2014-15	18.1
Bangladesh 2011	12.8	Sierra Leone 2013	18.1
Niger 2006	13.0	Afghanistan 2015	19.9
Chad 2004	13.2	Nigeria 2008	20.0
Ethiopia 2005	13.3	Sierra Leone 2008	21.1
Niger 2012	15.0		

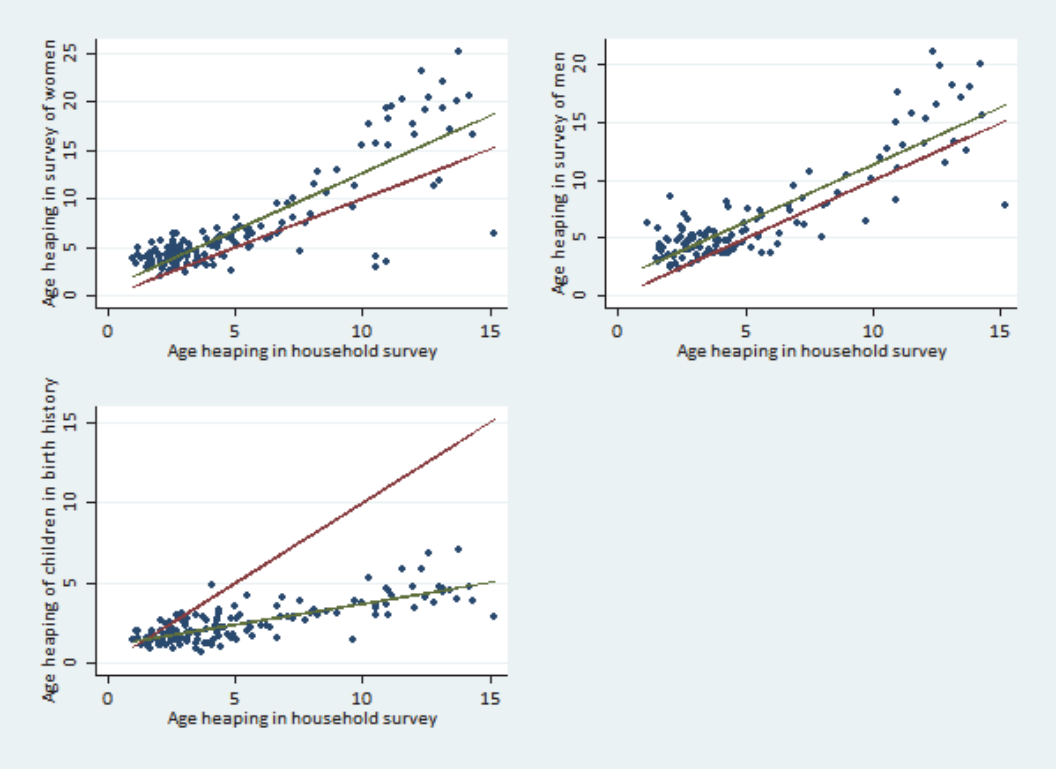
Finally, the mean level of heaping for children's ages in the birth histories is only 2.5%. No survey reaches a level of 10%. Only five surveys have Myers' Index values at or above 5%: Mali 2012-13 (5.3%), Sierra Leone 2008 (5.8%), Guinea 2005 (5.8%), Afghanistan 2015 (6.8%), and Chad 2014-15 (7.1%). These are all surveys that appear with high levels of the other heaping indicators.

A Myers' Index of 5% is very low. For example, compared with the expectation of 10% of cases at each final digit 0 through 9, the index would be 5% if we observed 13% at final digit 0, 12% at final digit 5, and less than 10% at all other final digits. With this hypothetical distribution, if the 3% excess at final digit 0

and the 2% excess at final digit 5 were redistributed to the final digits with deficits, then we could achieve a uniform distribution. The generally low value of this index for children ever born to the women in the survey of women, within the 30 years before the survey, may be due to the manner in which the birth history is constructed, starting from the first birth and working toward the most recent birth. The procedure places more emphasis on a birthdate, if it can be obtained, than on age. The children are positioned in a sequence. There may be other forms of age-related misreporting, such as tending to list children two years apart, but the age distribution of children is less heaped than the other age distributions.

Figure 2.4 shows the correspondences among the measures of age heaping. Age heaping in the household survey is the standard and is the horizontal axis in all three graphs. The red line represents hypothetical equality with the household survey's level of heaping. The green line is produced by regressing on the household survey's level of heaping.

Figure 2.4 Scatterplots showing the level of heaping in reported age in the survey of women, the survey of men, and the children in the birth histories, versus the heaping in reported age in the household survey



The upper left graph compares the age heaping in the survey of women with that in the household survey. The great majority of points are above the line of equality, showing that within each survey there tends to be worse heaping in the survey of women than in the household survey. As stated above, the difference is due to the different age ranges. The fitted line is above the line of equality, but roughly parallel to it, implying that the difference is mostly independent of the level. However, the correspondence is much closer when both levels are below a level of about 10% than when either of them is higher. There is considerable scatter when the levels are high. Four surveys have relatively high heaping in the household survey but relatively low heaping in the survey of women, and when the heaping in the household survey is in the range of 10%-15% there are about 20 surveys with heaping of women's ages in the range of 15%-25%. A similar pattern is seen in the upper right graph, for heaping of men's ages compared with the household survey, and in the lower left graph, for heaping of children's ages compared with the household survey. In

each graph, the point farthest to the right and farthest below the line of equality is for the India 2005-06 survey.

The pairwise correlations among the four heaping indicators are in a range from 0.79 to 0.91. The highest correlation (0.91) is between the women's index and the men's index. In all surveys, the interviewers are the same sex as the respondent. The high similarity in the extent of heaping—and of incompleteness, which was observed above—in the surveys of men and of women is probably due to the initial assessment of age during the household interview, in which the household interviewer and the household respondent determine (or negotiate) the ages of all the household members. Once those numbers are set, there is relatively little adjustment during subsequent interviews with the men and women who have been identified as eligible respondents for the surveys of women and men.

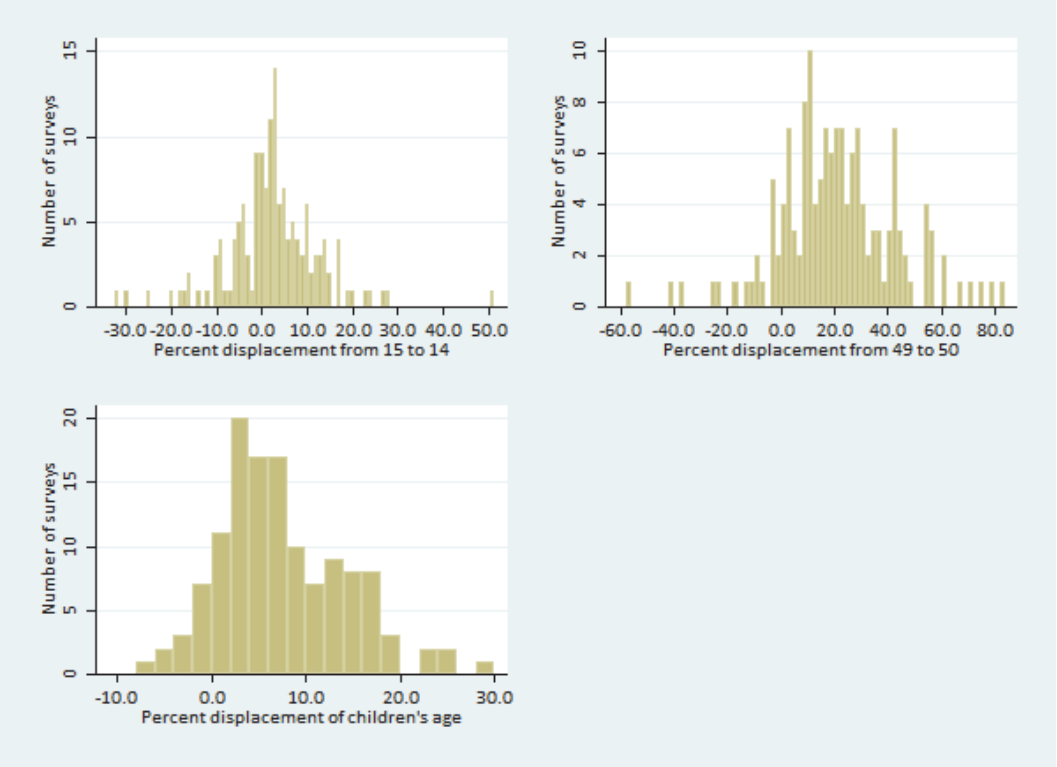
2.4 Displacement

Age displacement is an issue when a specific age range or time interval is used as a criterion for some purpose. In some situations, respondents may be motivated to adjust their age relative to an age of legal majority (e.g., 21) or military service (e.g., 18) or eligibility for retirement (e.g., 65). In the context of DHS surveys, it is the interviewer who may have an incentive to shift respondents out of an age range or time interval that is a criterion for eligibility for a longer interview. DHS has been sensitive to this phenomenon at age 14-15 and age 49-50 for women, and at the boundary for the health questions about children under age 5. Figure 2.5 shows the distributions of the displacement index for these three potential transfers. All of them are interpretable as percentages. For example, the index for displacement from age 15 to 14 is an estimate of the percentage of women with true age 15 who are misstated to be age 14. The procedure for calculating the "true" numbers at ages 15 and 14 was described earlier. Note that both the horizontal and vertical scales are different in the three graphs in Figure 2.5.

All three indicators are oriented such that a positive value is interpretable as a shift in the expected direction and a negative value as a shift in the opposite direction. A reference value of 0 means that there is no evidence of a shift—at least not with this particular indicator—but because of sampling variation we would expect some variation around 0 even if there were not a systematic tendency for cases to be shifted. In all three graphs, the great majority of transfers are in a positive direction, but some are in a negative direction (to repeat, the measure is aligned so that shifts in the expected direction are positive).

The average transfer from age 15 to 14 is only 2.6%, but with a very wide range, from -31.6% to 50.8%. Table 2.8 lists the surveys with the most extreme levels of transfers from age 15 to 14—including reverse transfers, indicated with negative values. These are all the transfers outside an arbitrary range from -15% to 15%, and they are accompanied by an approximate z statistic for a test of the null hypothesis that the population value was 0%, and by symbols to describe the level of significance.

Figure 2.5 Histograms showing the level of displacement from age 15 to 14, or from age 49 to 50, for women in the household survey, and displacement across the boundary for the health questions in the birth histories



In Tables 2.8 through 2.10, a "+" sign indicates significance in the expected direction and a "-" indicates significance in the opposite direction. One sign indicates significance with a two-tailed .05 test, two signs with a .01 test, and three signs with a .001 test. Adjustments for weights, clustering, and stratification are not included. Test statistics and p-values are included explicitly in the discussion of transfers because transfers can occur in either direction, and we emphasize that the listed values are not random variations around zero, but can be highly significant in either a positive (expected) or a negative (opposite) direction. In the lists of surveys with high levels of incompleteness or heaping, the measures were highly significant but test statistics and p-values were omitted.

Table 2.8 Surveys with age transfers from age 15 to 14 outside the range of -15% to +15%, and the statistical significance of those transfers

Survey	%	z	Sig.
Sierra Leone 2013	-31.6	-11.77	---
Nigeria 2013	-29.6	-15.58	---
Nigeria 2003	-25.0	-6.19	---
Ethiopia 2011	-19.7	-7.46	---
Nepal 2006	-18.2	-5.23	---
Cameroon 2011	-17.1	-5.91	---
Ethiopia 2000	-16.1	-5.97	---
Liberia 2013	-15.9	-4.26	---
Ukraine 2007	15.2	2.89	++
Myanmar 2015-16	16.9	4.76	+++
Namibia 2013	17.1	4.84	+++
Gabon 2012	17.2	4.55	+++
Azerbaijan 2006	17.4	4.33	+++
Sao Tome and Principe 2008-09	18.9	2.83	++
Swaziland 2006-07	19.6	4.62	+++
Congo Democratic Republic 2007	23.4	7.15	+++
Niger 2012	23.9	7.03	+++
Madagascar 2003-04	27.0	7.17	+++
Mali 2012-13	27.8	8.27	+++
Sierra Leone 2008	50.8	14.57	+++

The highest level of transfers in the expected direction, 50.8%, is found for the Sierra Leone 2008 survey. However, the highest level of transfers in the contrary direction, -31.6%, is found for the next survey in Sierra Leone, conducted in 2013. It is apparent that the high level in 2008 was noted, during or after fieldwork, and some procedures were adopted for the 2013 survey, through special training and supervision, in an attempt to prevent such transfers. Unfortunately, the net effect was an over-correction that moved older children into the time frame for the child health questions.

The Nigeria 2003 and 2013 surveys have the next-worst level of transfers from age 15 to 14 in a negative direction (that is, from age 14 to 15), almost certainly because the level of positive transfers was very high in the Nigeria 1999 survey (as noted in Pullum 2006) and an intentional effort was made to prevent such transfers, but the result was an over-correction. It is likely that other surveys with negative transfers also experienced over-correction for anticipated shifts from age 15 to 14.

In terms of relative frequency, the most serious type of transfer is from age 49 to 50, some of which is certainly due to heaping at age 50. Across all household surveys, the mean level is 21.3%, with a very wide range from -57.1% to 83.1%. Table 2.9 includes six surveys with negative shifts in excess of 15%. Four of the six surveys with extreme shifts in the unexpected direction, in effect from 50 to 49, were in Bangladesh—the surveys of 2004, 2007, 2011, and 2014. Clearly, these are the result of over-correction for an expected shift from age 49 to 50.

Many surveys had large shifts from age 49 to 50, so the threshold for identifying specific surveys will be set at a much higher level than 15%. Table 2.9 includes 14 surveys with shifts from 49 to 50 in excess of 50%. The list includes two surveys from Niger (2006 and 2012) and two surveys from Benin (2006 and 2011-12). It does not appear that there was any attempt to reverse the level of transfers following the first survey in each pair, because the second survey had the same or a higher level of transfers.

Table 2.9 Surveys with age transfers from age 49 to 50 outside the range from -15% to 50%, and the statistical significance of those transfers

Survey	%	z	Sig.
Bangladesh 2007	-57.1	-9.91	---
Bangladesh 2014	-41.8	-8.96	---
Bangladesh 2011	-36.9	-8.24	---
Rwanda 2014-15	-24.1	-4.10	---
Bangladesh 2004	-23.6	-4.03	---
Cambodia 2014	-17.7	-4.26	---
Niger 2012	54.2	9.95	+++
Niger 2006	54.2	8.44	+++
Benin 2006	54.2	13.24	+++
India 2005-06	55.3	31.14	+++
Gambia 2013	57.2	10.60	+++
Pakistan 2006-07	57.2	43.66	+++
Senegal 2014	57.3	8.22	+++
Angola 2015-16	60.0	14.17	+++
Senegal 2010-11	60.5	13.32	+++
Yemen 2013	66.5	18.77	+++
Comoros 2012	70.6	9.87	+++
Mali 2012-13	75.0	12.30	+++
Benin 2011-12	79.8	20.05	+++
Sierra Leone 2008	83.1	13.87	+++

Transfers across the boundary for the health questions have a mean value of 7.5%, with a range from -6.7% to 28.3%. The surveys listed in Table 2.10 had more than 15% displacement, all in the expected direction. The only countries appearing more than once are Malawi (2000 and 2004), Madagascar (2003-04 and 2008-09), and Egypt (2008 and 2014). Most of the countries on this list appear for early surveys within the time interval (2000-15) but are not repeated for later surveys. Both Malawi and Madagascar have had more recent surveys in which these kinds of transfers were less serious. It appears that this kind of transfer can be remediated without serious over-correction in the opposite direction.

Table 2.10 Surveys with average transfers for health that are 15% or greater and statistical significance of those transfers

Survey	%	z	Sig.
Sao Tome and Principe 2008-09	15.0	3.81	+++
Yemen 2013	15.2	11.43	+++
Egypt 2008	15.3	9.60	+++
Uganda 2006	15.8	8.67	+++
Swaziland 2006-07	16.1	5.01	+++
Egypt 2014	16.2	12.02	+++
Malawi 2004	16.6	9.87	+++
Ethiopia 2005	16.6	9.95	+++
Kenya 2008-09	17.1	8.04	+++
Madagascar 2003-04	17.2	7.60	+++
Madagascar 2008-09	17.7	12.17	+++
Afghanistan 2015	17.8	19.17	+++
Ghana 2008	18.3	6.00	+++
Benin 2006	18.4	13.72	+++
Indonesia 2002-03	18.5	13.87	+++
Malawi 2000	22.3	13.93	+++
Sierra Leone 2008	23.9	10.96	+++
Mali 2006	24.9	17.99	+++
Niger 2006	25.1	14.77	+++
Niger 2012	28.3	19.33	+++

Figure 2.6 Scatterplot showing the level of displacement from age 15 to 14 versus the level of displacement from age 49 to 50, for women in the household survey

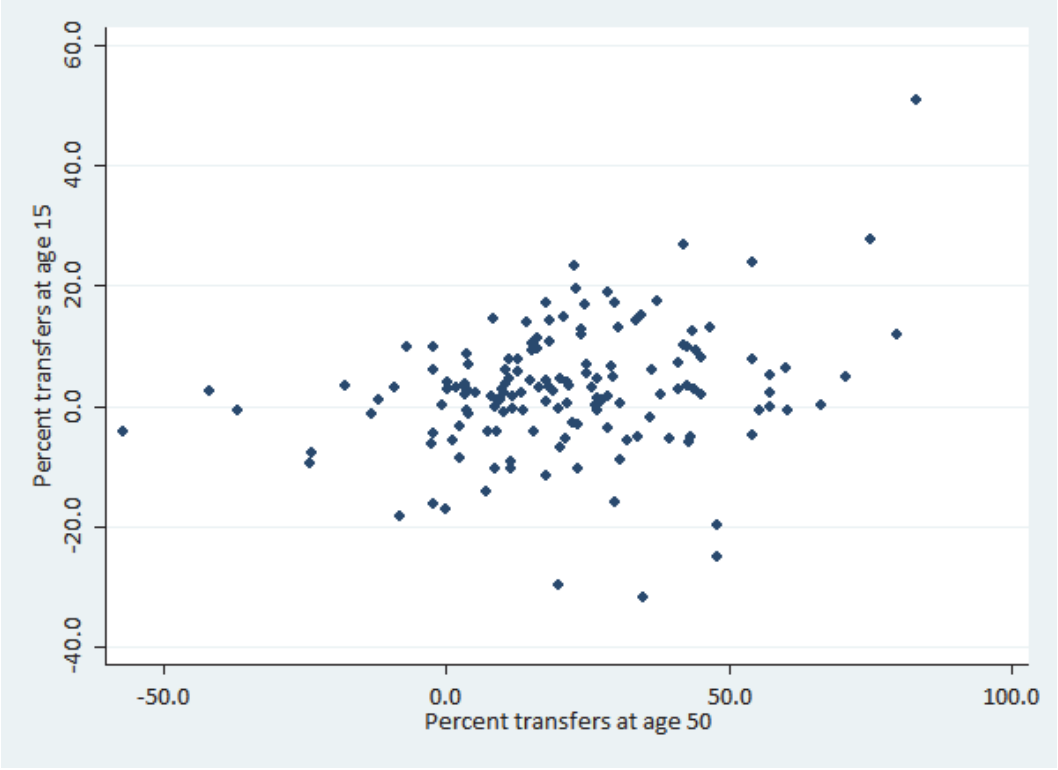


Figure 2.6 gives a scatterplot of the two types of shifts in the in the household age distribution. There is no evidence of statistical association between them (there is also no evidence of association with transfers of children). However, the Sierra Leone 2008 survey had the highest level of transfers from age 49 to 50 (83.1%) and the highest level of transfers from age 14 to 15 (50.8%). It also had the second highest level of transfers of children across the boundary for the health questions (23.9%), second only to Niger 2012 (28.3%). In the Sierra Leone 2013 survey, the level of transfers from age 14 to 15 was reversed because of over-correction of the direction of transfers in the earlier survey; the other two transfers remained in the expected direction but were more moderate.

2.5 Composite Measures of Data Quality

Finally, we attempt to summarize changes from 2000 to 2015 by constructing standard scores for the three types of misreporting, with the following three steps:

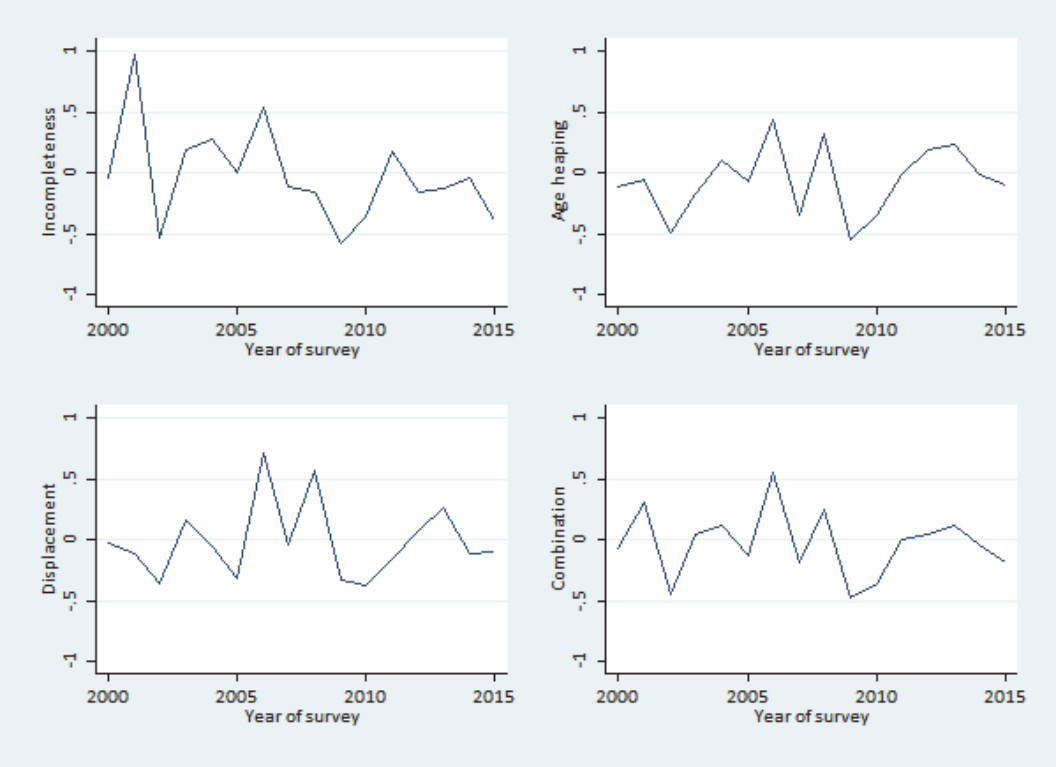
First, the displacement indicators are replaced with their absolute values. The incompleteness and heaping indicators all have natural zeros—they can never be negative, and under ideal conditions they would all be zero except for sampling error—but the displacement indicators can be positive or negative. They were constructed to be positive for displacement in the expected direction, but they can be negative if, say, there was over-correction. The absolute value measures the magnitude of displacement, regardless of the direction.

Second, the mean and standard deviation of each indicator were calculated, and then a standard score (or z-score) was calculated by subtracting the mean and dividing by the standard deviation. That is, each indicator is represented by a new score that has a mean of 0 and a standard deviation of 1. By doing this, we put all 11 indicators, which vary widely in their means and standard deviations, onto the same scale.

Third, we calculate the average of the z-scores within each group. “Incompleteness” is the average of the four z-scores in that group of indicators, “Age heaping” is the average of the four z-scores in that group, and “Displacement” is the average of the three z-scores in that group. In each case, the average is based on the number of indicators of each type that are available. “Combination” is an average of the z-scores for all 11 indicators, giving the same weight to all of them.

The purpose of Figure 2.7 is to see how each of these averages has changed over time, within the interval 2000 to 2015. If, say, there were clear and steady improvement in the level of incompleteness, then the upper left graph would start well above 0 in 2000 and steadily move downward, ending well below 0 in 2015. For all of the figures, a downward trend would be interpreted as improvement in the quality of the age and date reporting. Each indicator measures an undesirable characteristic, so low values are preferable to high values.

Figure 2.7 Changes in the levels of incompleteness, heaping, displacement, and a combination of all three in DHS surveys conducted from 2000 to 2015



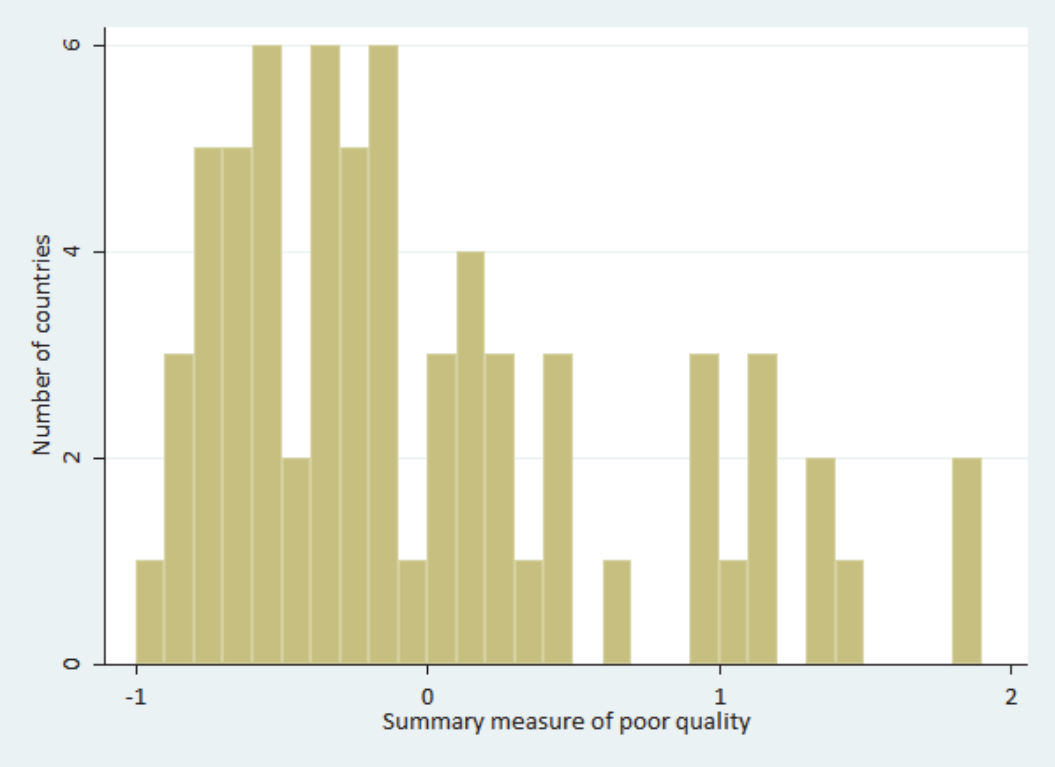
All four lines show some decline during the final three years, 2013-2015, but for the full interval 2000-2015 there is no convincing evidence of improvement. All the figures show an upward spike in 2006, to the maximum values for incompleteness, heaping, and displacement, with the exception of the spike for incompleteness in 2001.[9] Each type had a minimum in 2002, moved upward to 2006-2008, had another minimum in 2009-2010, moved upward to 2013, and then moved back down to the most recent year, 2015. Some short-term variation is clear, but, to repeat, there is no clear evidence of long-term improvement between 2000 and 2015.

The combination indicator, which averages the 11 z-scores for the full set of indicators, varies considerably across surveys.[10] In Figure 2.8, we take the indicator one step further by calculating an average of the combination indicator for all the surveys that have been conducted in a given country. This step allows us to rank countries in the quality of their data collected during 2000-2015. It should be helpful, when planning another survey in a country where DHS has worked before, if DHS has objective evidence of past problems and can target training and resources to avoid a recurrence of those problems.

[9] There might appear to be a downward trend with incompleteness, which would continue an earlier pattern noted in Pullum (2006), but this impression is simply due to the spike in incompleteness in 2001.

[10] The Combination indicator is equivalent to the first principal component of a principal component analysis of the 11 indicators.

Figure 2.8 Distribution of the summary measure of incompleteness, heaping, and displacement across the 148 surveys in this analysis



In Figure 2.8, lower values on the horizontal axis indicate better quality data. The reference value of 0 is the mean for all 148 surveys, and is approximately the mean for all 67 countries represented in the figure. The distribution is skewed to the right, with 13 countries having a separation from the main body of the figure and a combination level of 0.6 or above. It is somewhat arbitrary where to draw a boundary, but these countries represent the bottom quintile in data quality. We list them in Table 2.11, ordered by the combination level but including the standard scores for incompleteness, heaping, and displacement, to give a clearer idea of what dimension(s) of age and date reporting are most problematic. Sierra Leone and Niger have virtually identical combination scores, the highest on the list. Most countries in the table are in sub-Saharan Africa, with the exceptions of Afghanistan, Bangladesh, Pakistan, and Yemen.

Table 2.11 Countries in the poorest quintile of data quality for ages and dates, based on a combination of standardized scores for incompleteness, heaping, and displacement in DHS surveys during 2000-2015

Country	Surveys	Incomp.	Heaping	Displacement	Combination
Comoros	1	0.53	0.65	0.72	0.63
Pakistan	2	1.31	1.67	0.02	0.95
Guinea	2	0.83	2.01	-0.01	0.98
Bangladesh	4	2.23	1.01	-0.13	1.00
Nigeria	3	-0.05	2.11	1.03	1.01
Mali	3	1.08	1.02	1.26	1.12
Yemen	1	1.64	1.11	0.83	1.14
Ethiopia	3	1.55	1.36	0.46	1.16
Benin	3	1.60	1.43	0.94	1.32
Chad	2	1.30	2.28	0.18	1.39
Afghanistan	1	0.58	2.82	0.96	1.45
Niger	2	2.01	1.59	1.86	1.82
Sierra Leone	2	0.64	2.39	2.47	1.83

The 13 countries comprising the lowest quintile of the combination score, that is, the best quintile of data quality, are listed in Table 2.12. Guatemala, which had only one DHS survey (in 2014-15), has the lowest combination score of all surveys. This list includes other countries in Latin America and the Caribbean, namely Colombia, Guyana, Honduras, and Peru. It also includes Albania, Armenia, Kyrgyzstan, Moldova, and Tajikistan, in Eastern Europe and Central Asia, and one country in Southeast Asia—the Philippines—and one in sub-Saharan Africa—Zimbabwe.

Of the countries in Table 2.11, there are three where DHS conducted only one survey within the time interval. In Table 2.12, there are seven where DHS conducted only one survey within the time interval. However, it cannot be inferred that the quality of age and date reporting has nothing to do with accumulated experience, because many of the countries in the best quintile had DHS surveys prior to 2000.

The identification of specific surveys and countries having relatively poor data, or relatively good data, in terms of age and date reporting, is based on diagnostic tools that are limited in their reliability and validity and are themselves subject to sampling and measurement error. Good quality and poor quality have many potential sources, of which one of the most important is simply the relevance of knowing one's age or birthdate with any accuracy in the context of daily life in each country. If respondents do not know their age, because they do not need to know their age for practical purposes, for example, then no amount of training of interviewers can substitute. Nevertheless, knowledge of the surveys in which the reporting of ages and dates has had a relatively high frequency of problems may be useful for redirecting resources in anticipation of potential problems in future surveys.

Table 2.12 Countries in the highest quintile of data quality for ages and dates, based on a combination of standardized scores for incompleteness, heaping, and displacement in DHS surveys during 2000-2015

Country	Surveys	Incomp.	Heaping	Displacement	Combination
Guatemala	1	-1.04	-0.87	-0.90	-0.93
Colombia	4	-0.83	-0.89	-0.80	-0.83
Jordan	3	-0.94	-0.94	-0.68	-0.83
Philippines	3	-0.95	-0.92	-0.57	-0.80
Honduras	2	-0.74	-0.80	-0.75	-0.76
Albania	1	-1.02	-0.67	-0.57	-0.75
Peru	6	-0.79	-0.67	-0.73	-0.73
Republic of Moldova	1	-0.97	-0.54	-0.64	-0.72
Kyrgyzstan	1	-1.03	-0.51	-0.60	-0.71
Guyana	1	-0.59	-0.81	-0.68	-0.69
Armenia	3	-1.04	-0.70	-0.30	-0.68
Tajikistan	1	-1.08	-0.74	-0.28	-0.64
Zimbabwe	3	-0.92	-0.73	-0.23	-0.63
Viet Nam	1	-0.99	-0.52	-0.45	-0.62

3 The Quality and Consistency of Spousal Age Estimates in DHS Surveys

Core DHS questionnaires for women first began to include questions about age of husband or live-in partner in DHS Phase 3 (Macro International 1995a, 1995b). Since then, in most DHS surveys women who are married or living with a man as if married are asked the age of their partner. Core questionnaires for men were first included in DHS core questionnaires in Phase 5 (Macro International 2008b); men who are married or living with a woman or with multiple women as if married are asked the age of their partner(s). These questions facilitate an analysis of their partner's age and their relative age difference in relation to demographic and health outcomes even in cases where the partner is not interviewed or is not a member of the household.

In surveys and households where men are interviewed, it is possible to compare women's estimates of their husband's or partner's (herein: spouse's) age and men's self-reported age. Additionally, in many recent surveys where men are asked the age of their spouse(s) a similar comparison between women's self-reported age and their spouse's estimate of their age is possible.

While it may be presumed that self-reported ages are more accurate than spousal estimates of age, when the two reports conflict it is unknown which—if either—reflects the 'true' value of age. Instead, we can examine consistency and heaping of self-reports and spousal estimates of age. In this section we explore the composition of partnerships by self-reported age difference and the consistency of spousal estimates and self-reports of age. Additionally, we test four hypotheses about spousal versus self-reported age:

- Hypothesis 1: The age difference between partners tends to be reduced by spousal estimates.
- Hypothesis 2: Spousal age estimates have more heaping than self-reported age.
- Hypothesis 3: The older the individual, the less consistent estimates of their spouse's age will be with self-reported age.
- Hypothesis 4: Individuals with primary, secondary, and higher education will provide more consistent estimates of their spouse's age than individuals with no education.

3.1 Data and Methods

Data for this analysis come from DHS couples' files, which include all couples where both the man and woman were living in the same household, were individually interviewed by DHS, and each identified the other as someone they were married to or living together with as if married. Women, who are never asked by DHS surveys about multiple husbands, may appear in the couples file only once. Men married to more than one woman who was interviewed by DHS will appear in each applicable couple in the dataset.

Surveys used in this analysis are Standard and Continuous DHS surveys fielded in 2000 or after where men were interviewed and where women and/or men were asked about their spouse's age. In total, 113 surveys qualified for inclusion. Appendix D shows a list of these surveys. As the table indicates, women were asked about their spouse's age in all of these surveys, but men were asked about their spouse's age in only 67 surveys. All analysis is unweighted, both at the survey and couple level.

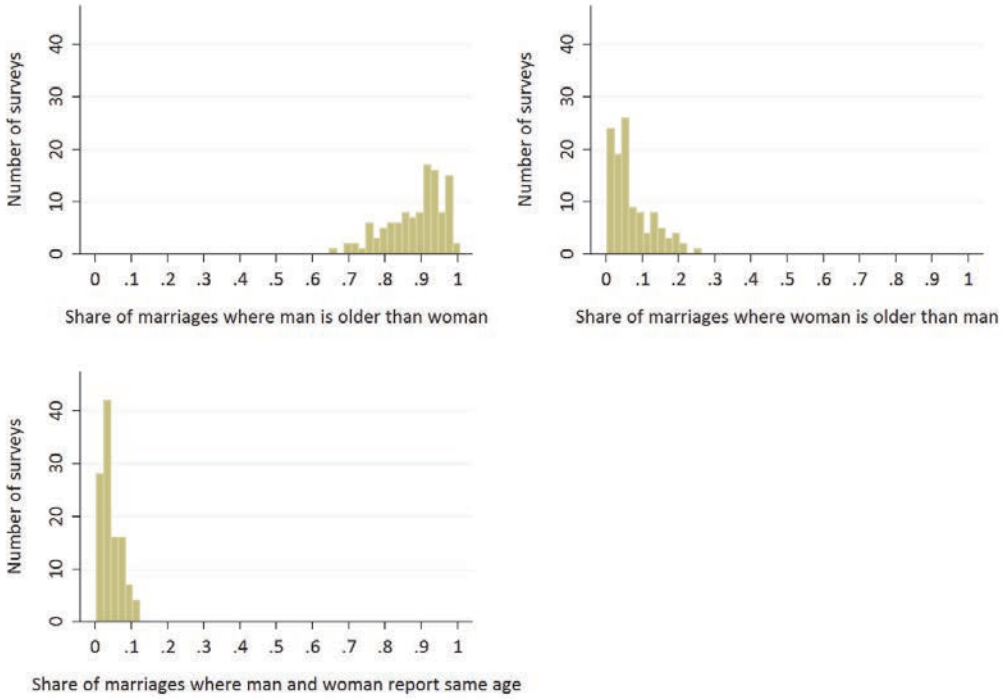
We analyze the composition of couple's age differences, the direction of differences of spousal estimates, heaping in spousal estimates, and consistency of estimates by age and education. For tests of heaping (Myers' Index), all ages had to be restricted such that they would be allowed to span an equal number of ages ending in digits 0 to 9. The minimum and maximum age bounds for heaping analysis and for all other

analysis are shown in Appendix D. Note that in 16 surveys men's age was allowed to be as low as 20, but in the couple's file men's minimum age was actually 21 or 22. For consistency of estimates by age and education, regression analysis was run to predict the absolute and percentage difference of spousal age estimates based on the age or education of the husband or wife who was assessing their partner's age.

3.2 Composition of Couples by Direction of Age Difference

In each of the 113 surveys included in this study, for marriages and live-in partnerships (herein marriages) where both members of a co-resident couple were individually interviewed and where each member reported the other as a spouse or someone with whom the respondent was living with as if married, we examined the responses for differences between the two members in self-reported ages. Figure 3.1 shows the distribution of the share (proportion) of marriages by direction of age difference in all 113 surveys. The share of marriages where the man is older than the woman ranges from 65% in Myanmar (2015-16) to 99% in Guinea (2012), Mali (2006), and Niger (2006, 2012), with an average of 89% and a median of 91% across surveys. The distribution of the percentage of marriages where the woman's self-reported age is greater than the man's self-reported age ranges from <0.01% in Niger (2006, 2012) to 25% in Myanmar (2015-16), with a survey average of 7% and a survey median of 5%. The share of marriages where the man's and woman's self-reported ages are equal averages and has a median of 4% across surveys, and ranges from 0.04% in Niger (2012) to 11% in Bolivia (2003), Moldova (2005), Myanmar (2015-16), and Ukraine (2007).

Figure 3.1 Distribution of the share of marriages by direction of age difference



3.3 Difference between Spousal Age Estimates and Self-Reports

On average across surveys, the difference between women's estimates of their husband's age and husband's self-reports is -0.03 years, and the average absolute difference is 0.77 years. The cross-survey average difference between men's estimates of their wife's age and wife's self-reports is -0.15 years and the average absolute difference is 0.53 years. These near-zero averages disguise important variation; the average survey standard deviation of the difference between women's estimates of their husband's age and husband's self-

reports of their age is 2.21 years, and the average survey standard deviation of the difference between men's estimates of their wife's age and wife's self-reports is 1.46 years. These standard errors suggest some important differences between spousal estimates and self-reports of age.

In this section we test the first of our four hypotheses about spousal age estimates—that the age difference between partners tends to be reduced by spousal estimates. To test this hypothesis we examine, for each of the differential assessments—wife's estimate of husband's age versus husband's self-report, and husband's estimate of wife's age versus wife's self-report—the distribution of the difference across three types of marital age differences: man is older, woman is older, and man and woman are the same age. Note that these classifications of marital age difference are themselves based on self-reported ages.

Figure 3.2 shows the distribution of the survey average difference between wife's estimate of husband's age and husband's self-reported age in the three types of marriages. In marriages where the man is older, the average of this difference across surveys ranges from -1.03 to 0.59 years, with a mean of -0.11; in other words, women whose husbands are older tend to estimate that their husbands are slightly younger than their self-reported age. This lends support to the hypothesis that spousal age estimates tend to reduce the age difference between spouses.

Figure 3.2 also shows the distribution of the same survey metric—average difference between wife's estimate of husband's age and husband's self-reported age—is shown for marriages where the woman is older. Here we see nearly universal support for the hypothesis that women's estimates of their husband's age tends to reduce the age difference between them: women who are older than their husband tend to estimate their husband's age as older than their husband's self-reported age. The survey average of this age difference ranges from -0.03 to 6.86 years, with a cross-survey average of 1.20 years older.

The distribution of the average difference between wife's estimate of husband's age and husband's self-reported age is also shown in Figure 3.2 for the baseline group: marriages where the woman and man self-report the same age. Here, women tend to estimate their husband as older than the husband's self-report, with an average of 0.70 years across surveys, and a range from -0.06 to 3.37 years. This makes the earlier result for marriages where the woman is older appear more normative, although still greater than baseline. It also makes the result for marriages where the man is older appears more impressive: women may normally tend to estimate their husbands as older than husband's self-reports, but in marriages where the men are older, women estimate them to be younger than men's self-reports. Thus in marriages where a spousal age difference exists, the evidence is consistent with the hypothesis that women's estimates of their spouses ages tend to reduce the age difference between them; in the absence of a spousal age difference, women tend to estimate their husbands to be older than husbands self-report.

Figure 3.2 Distribution of average difference between wife's estimate of husband's age and husband's self-reported age, by marriage type

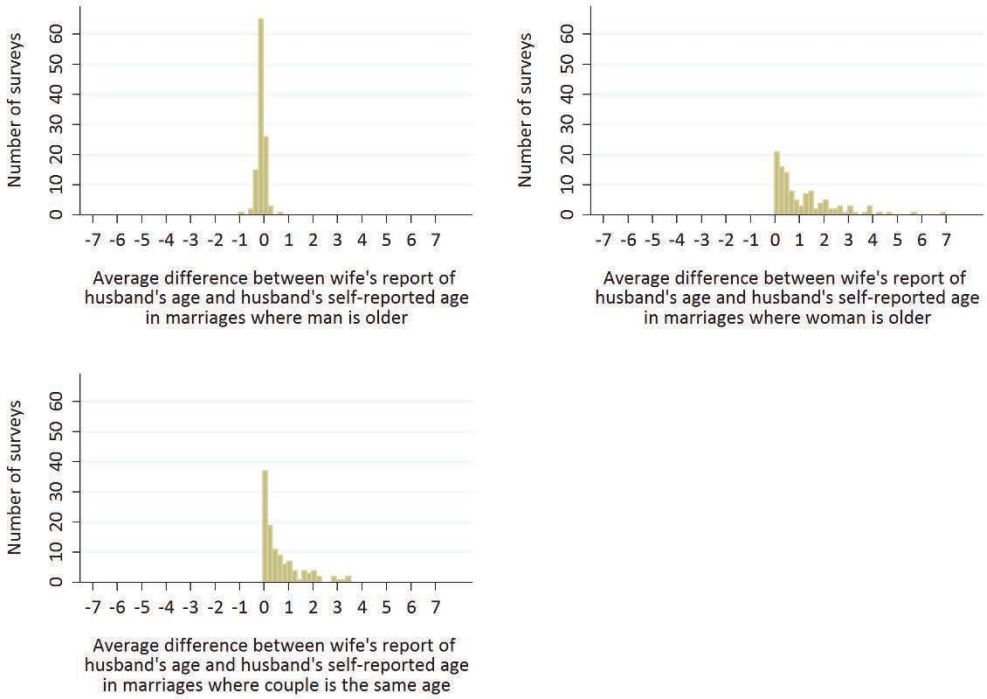
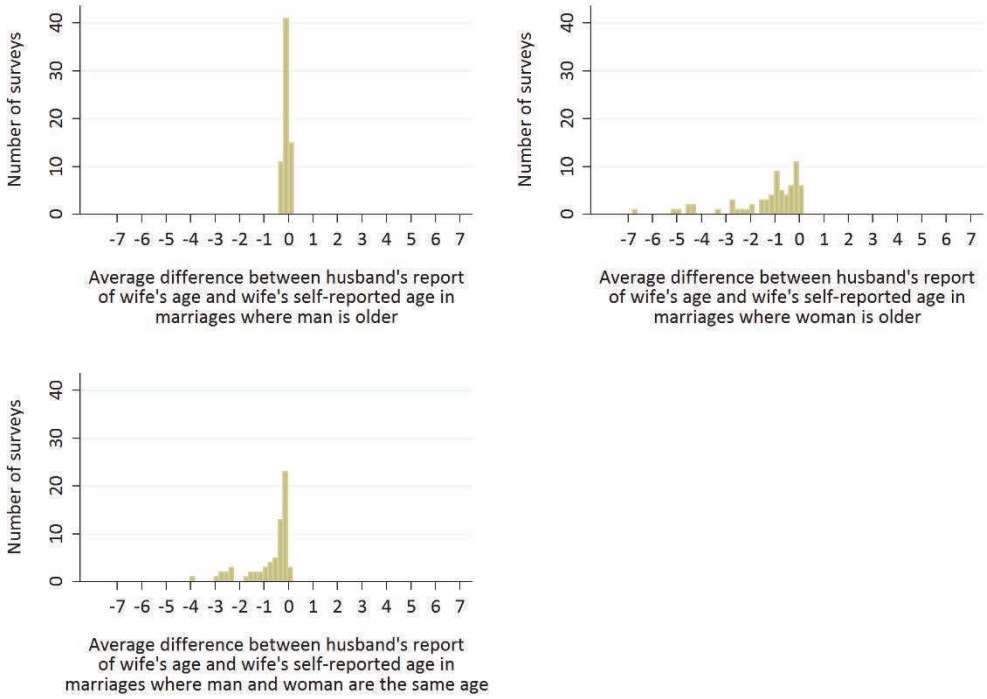


Figure 3.3 shows the results of our test of the same hypothesis on men's estimates of their wife's age. The average difference across surveys between husband's estimate of wife's age and wife's self-reported age for marriages where the man is older ranges from -0.40 to 0.10 years, and averages -0.08 years. This is generally inconsistent with our hypothesis that spousal estimates tend to reduce the age difference: men who self-report as older than their wife tend to slightly underestimate their wife's age. In the distribution of the same survey metric for marriages where the woman self-reports as older, we see a near-universal average of less than zero, meaning men tend to underestimate their wife's age when their wife self-reports as older. The average of this metric across surveys ranges from -6.83 years to 0.05 years, with an average of -1.31 years. Finally, the distribution of the average difference across surveys between husband's estimate of wife's age and wife's self-reported age in the baseline group—marriages where both partners self-report the same age—ranges from -4.04 to 0.04 years, with an average of -0.74. These results suggest that men generally tend to underestimate their wife's ages, more so if the wife self-reports to be older and less so if the wife self-reports to be younger than the husband.

Figure 3.3 Distribution of average difference between husband's estimate of wife's age and wife's self-reported age, by marriage type



3.4 Heaping in Spousal Age Estimates

In this section we test our second hypothesis—that spousal age estimates have more heaping than self-reported age. As in the earlier chapter, we compare the Myers' Index estimate of heaping, which can be interpreted as the minimum percentage of cases that would have to be shifted from one digit to another in order to achieve a uniform distribution across ages.

Figures 3.4a and 3.4b show a comparison of the Myers' Index for self-reported ages and spousal estimates of age. In six surveys, all from West Africa, Myers' Index for self-reported ages among men is above 20%—Nigeria 2008, Sierra Leone 2008, Nigeria 2013, Benin 2006, Nigeria 2003, Sierra Leone 2013—and in all of these cases Myers' Index is higher for spousal estimates than for self-reports. In five surveys—Dominican Republic 2002, Indonesia 2012, Zambia 2013-14, Haiti 2012, and Dominican Republic 2007—Myers' Index is below 3%, and in all except Gabon Myers' Index is higher for spousal estimates. In all, the correlation coefficient for Myers' Index of self-reported age and Myers' Index of spousal estimates of age is 87%. In 91 of 113 surveys (81%) Myers' Index is higher for spousal estimates than for self-reports of age. These results for men support the hypothesis that there is more heaping in spousal estimates of age than in self-reports.

To compare heaping of spousal age estimates versus self-reported age among women, we computed the Myers' Index for all 67 surveys in which men were asked to estimate their wife's age. Figure 3.7 shows a comparison of Myers' Index for self-reported ages among women and Myers' Index for spousal estimates of age. The estimates for self-reported age range from 1% in Myanmar 2015-16 to 24% in Sierra Leone 2008. In 53 of 67 surveys (79%), Myers' Index is higher for spousal estimates of age than for self-reports. Hence, the results for women also support the hypothesis that there is more heaping in spousal estimates of age than in self-reports.

Figure 3.4a Myers' Index for men's ages, self-reports versus spousal estimates

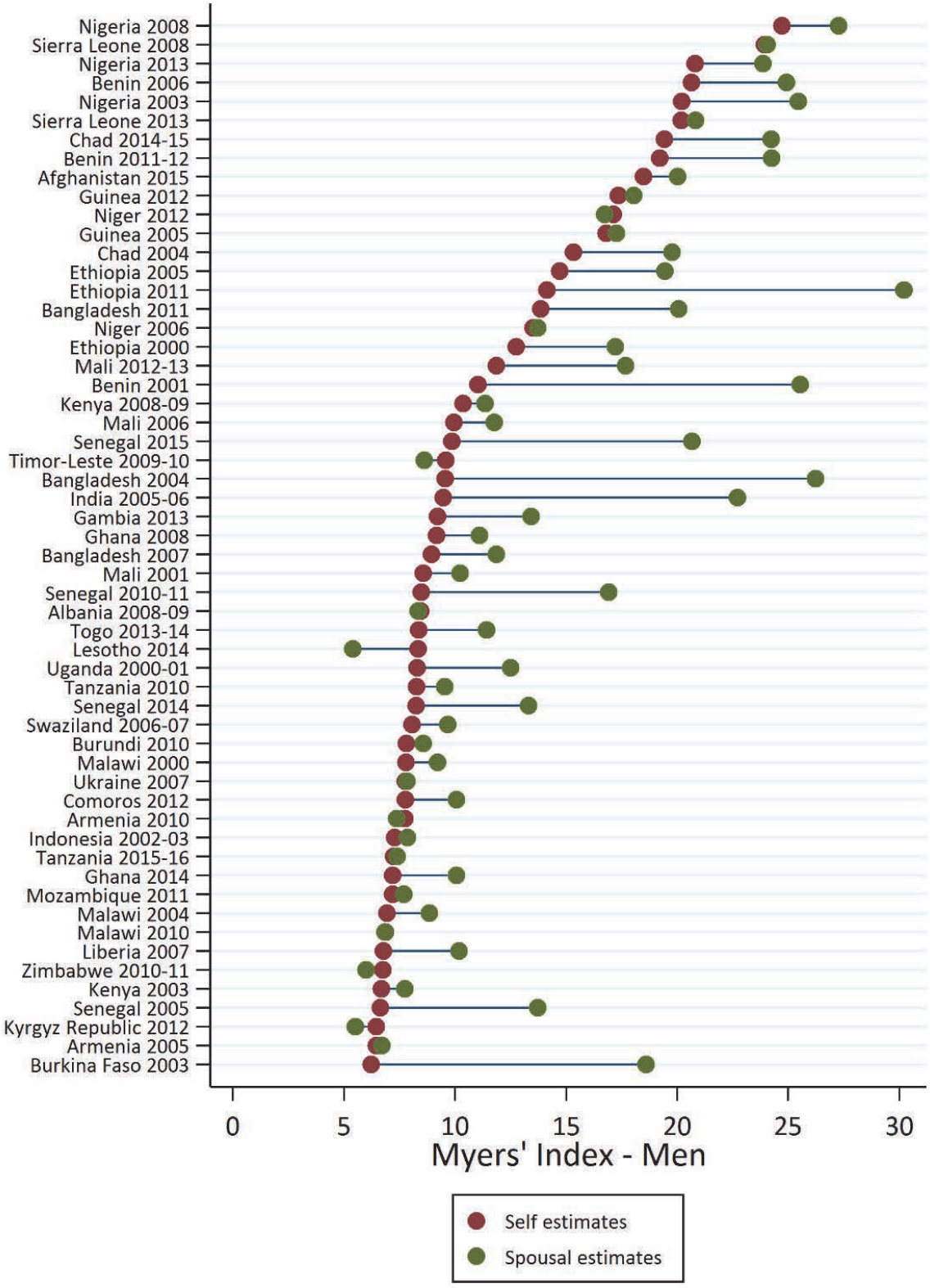


Figure 3.4b Myers' Index for men's ages, self-reports versus spousal estimates (continued)

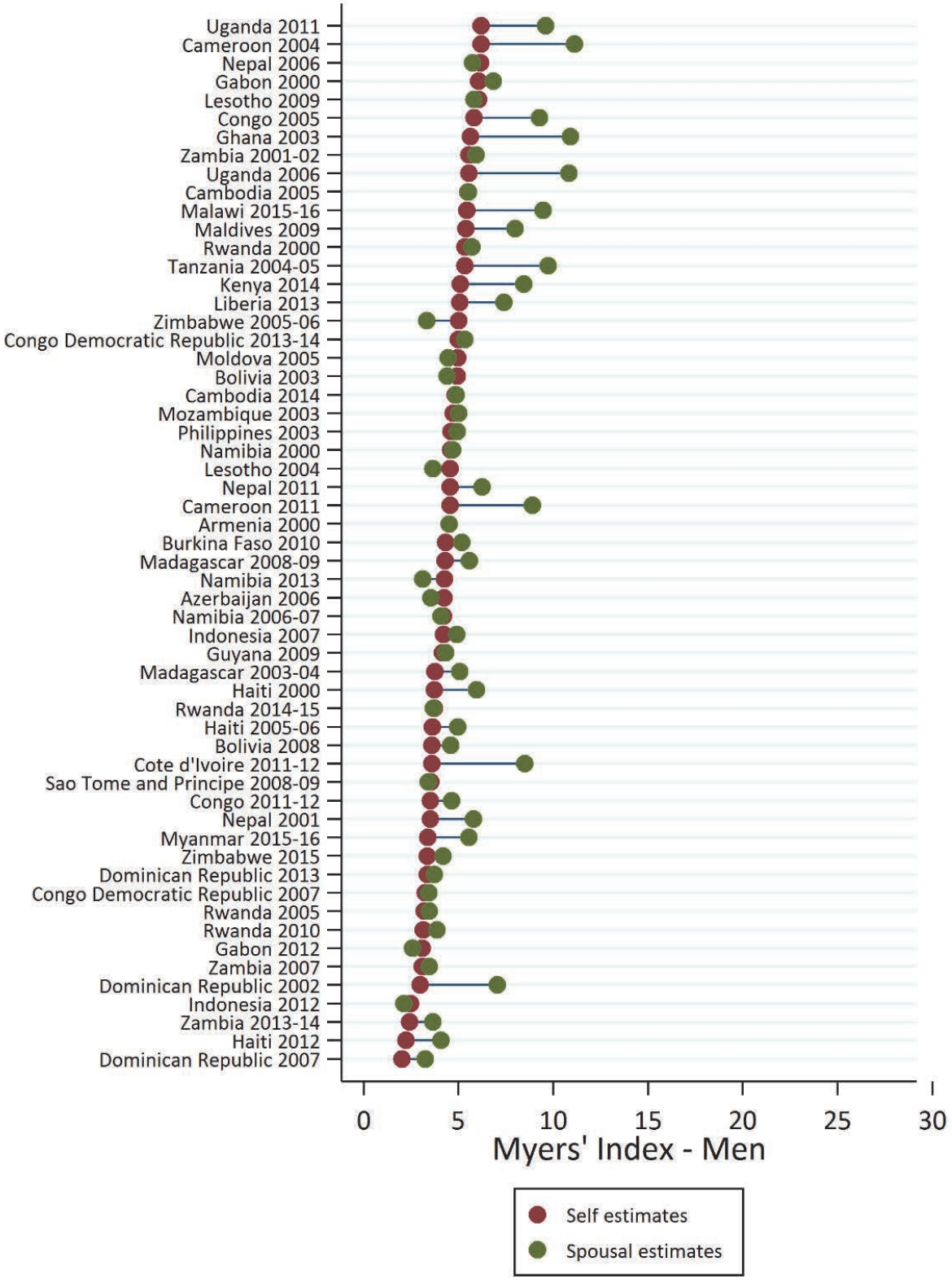
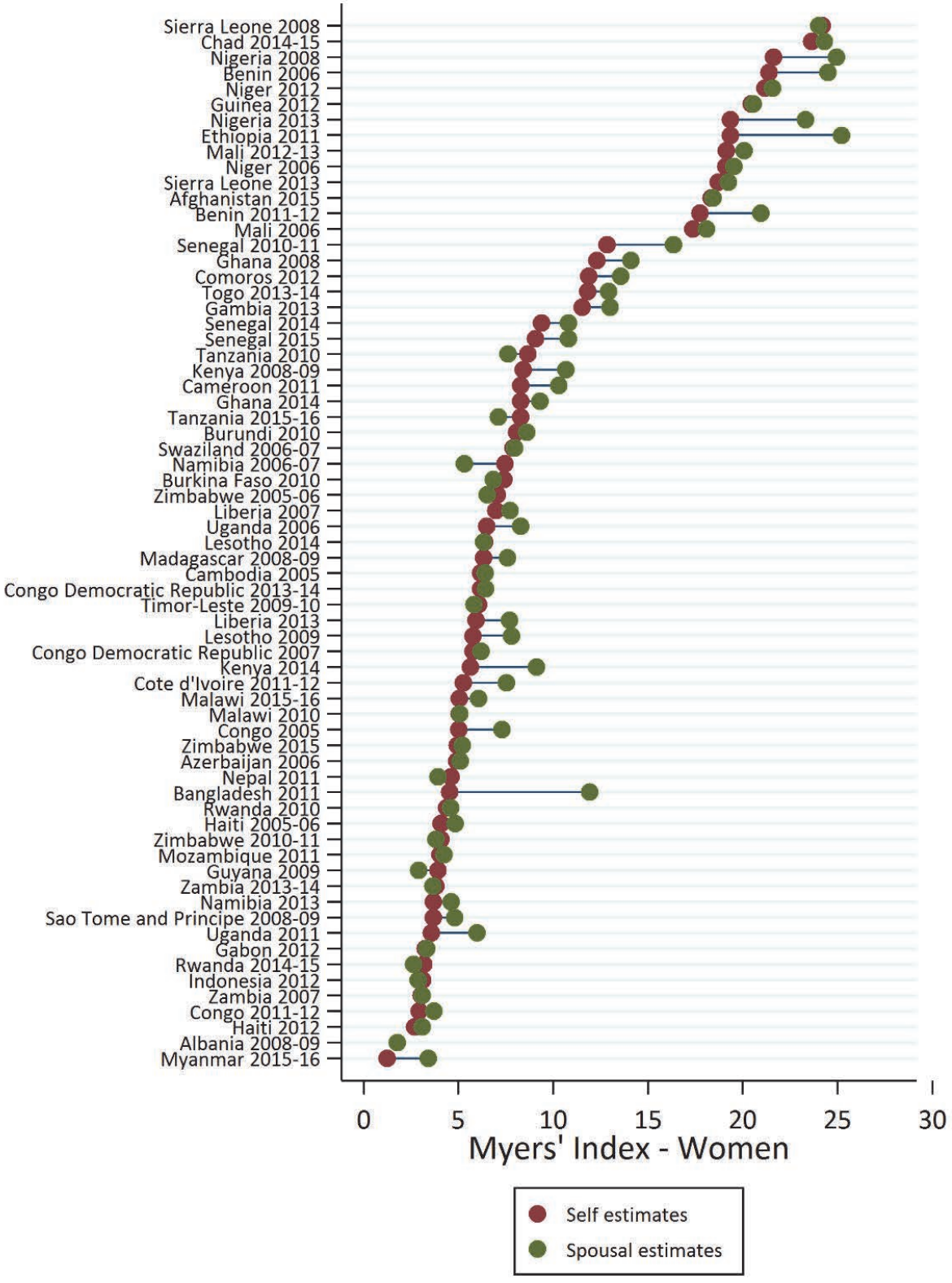


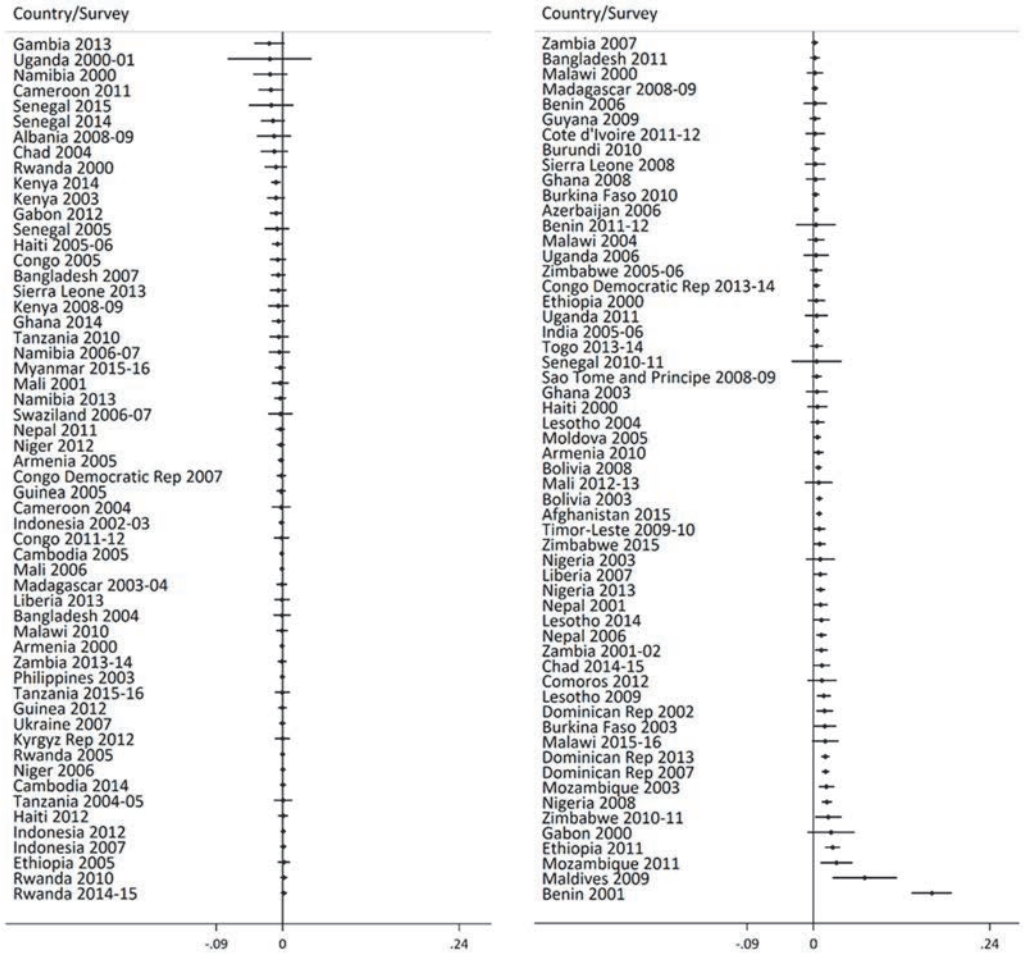
Figure 3.5 Myers' Index for women's ages, self-reports versus spousal estimates



3.5 Consistency of Spousal Estimates by Age

Here we test our third hypothesis—that the older the individual, the less consistent estimates of their spouse's age will be with self-reported age in absolute terms. To test this hypothesis we ran unweighted regressions of the absolute difference between spousal estimates and self-reported age by age of spouse for each survey. Figure 3.6 shows the coefficients of woman's self-reported age on the absolute difference between women's estimates of their husband's age and husband's self-reported age, in ascending order of coefficient. In 45 of 113 surveys the coefficient is negative, but in only one survey (Kenya 2014) is it significantly less than zero. This indicates that in Kenya the lower the wife's age the less consistent her estimation of her husband's age is with his own reported age. In 21 of the 68 surveys where the coefficient on wife's age is positive it is also statistically significant—that is, the higher the wife's age the less consistent her estimate of her husband's age is with the husband's own report. Notably, in the Maldives 2009 and in Benin 2001 the coefficient is significant and quite high in relative terms (0.07 and 0.16, respectively), indicating for example in Benin 2001 that for each additional year of wife's age the absolute difference between her estimate of her husband's age and his self-reported age is predicted to increase by 0.16 years. These results provide support for our hypothesis in almost one-fifth of the surveys studied, and provide contrary evidence in just 1 of the 113 surveys studied.

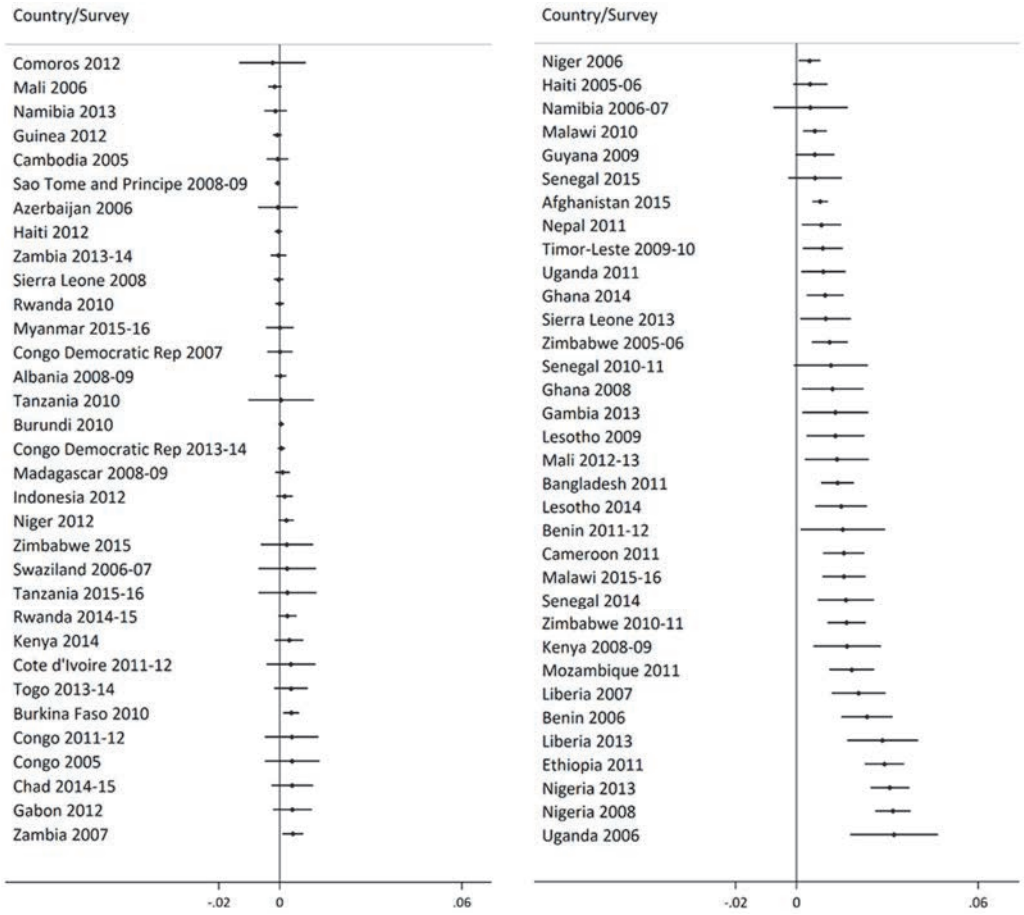
Figure 3.6 Coefficient for wife's self-reported age from regression on absolute difference between husband's self-reported age and wife's estimate of husband's age



We also tested the age hypothesis on husband's estimates of wife's age. The results are shown in Figure 3.7, which—as with subsequent charts—uses a different x-axis scale. As Figure 3.7 shows, in 11 of 67

surveys the coefficient on women's age is negative, but not significant. In 31 of the remaining 56 surveys where the coefficient is in the expected direction (positive), it is also statistically significant. Notably, the coefficients never reach even half of the level they did in Maldives 2009 or Benin 2001 for men, which suggests that while age may be statistically significant it is less salient of a factor in accuracy. Taken together, these findings provide modest support for the hypothesis that the absolute value of differences between spousal estimates and self-reports of age is higher at older ages.

Figure 3.7 Coefficient for husband's self-reported age from regression on absolute difference between wife's self-reported age and husband's estimate of wife's age



3.6 Consistency of Spousal Age Estimates by Education

To test our fourth hypothesis—that individuals with primary or secondary-plus education will provide more consistent estimates of their spouse's age than individuals with no education—we ran unweighted regressions of the absolute difference between spousal estimates and self-reported age by education of spouse (primary versus none, secondary and higher versus none) for each survey. Armenia 2005 and Ukraine 2007 were excluded from this analysis because in these two surveys all women in the couples file had some education.

Figure 3.8 shows the coefficients on wife's primary education (versus none) on the absolute difference of her estimate of husband's age and his self-reported age, in ascending order of coefficient size. In nearly half of the surveys the coefficient is negative, and in 22 surveys it is negative and statistically significant, consistent with our hypothesis. The significant effect of primary education on accuracy was largest in Senegal 2010-11, Zimbabwe 2010-11, Benin 2001, Senegal 2015, Cameroon 2011, and Dominican

Republic 2002. However, in six surveys—Mali 2006, Zambia 2007, Burkina Faso 2010, Timor-Leste 2009-10, Afghanistan 2015, and Kenya 2014—the coefficient is positive and statistically significant, counter to our hypothesis.

Figure 3.8 Coefficient for wife having primary education (versus none) from regression on absolute difference between husband's self-reported age and wife's estimate of husband's age

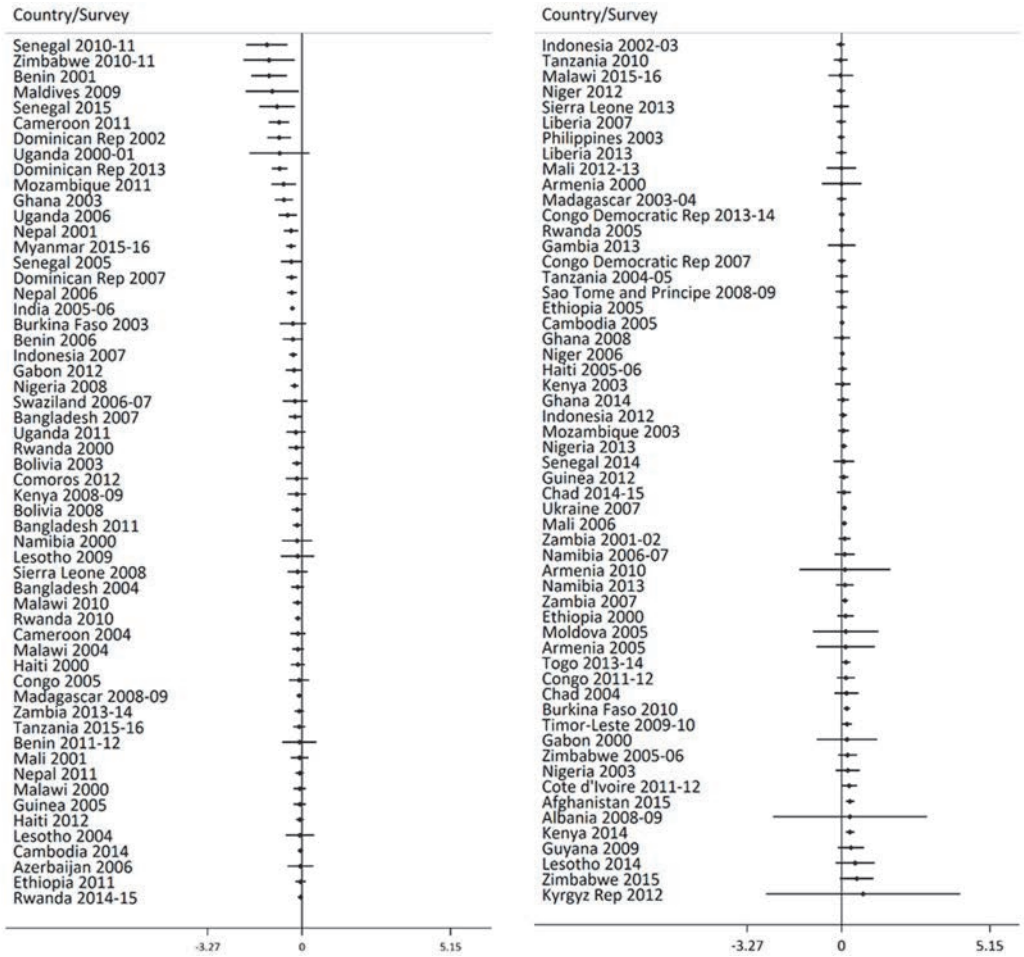
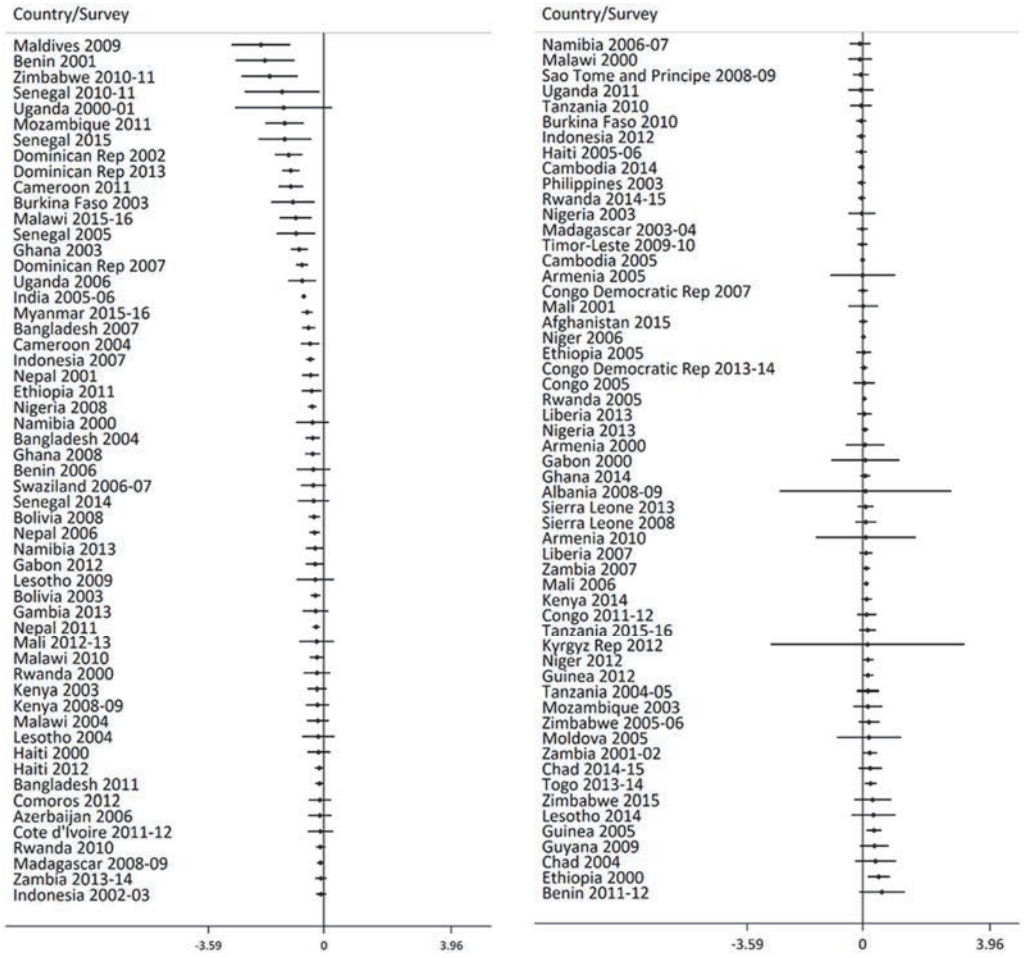


Figure 3.9 shows coefficients on wife's secondary education (versus no education) from the same regression, in ascending order. Here, we would expect the effect of education on consistency of ages to be stronger, and in general it is, both in magnitude and in frequency. In 70 of 111 surveys, the coefficient on secondary education is negative, and in 29 cases the coefficient is both negative and statistically significant, which supports our hypothesis. The significant effect of secondary education on accuracy is strongest in Maldives 2009, Benin 2001, Zimbabwe 2010-11, and Senegal 2010-11. Meanwhile, in six cases the coefficient is positive and statistically significant, counter to our hypothesis.

Figure 3.9 Coefficient for wife having secondary or higher education (versus none) from regression on absolute difference between husband's self-reported age and wife's estimate of husband's age



The same regression model on the absolute difference between spousal estimates and self-reported age was run for husband's education (primary versus none and secondary or higher versus none). Figure 3.10 shows results for the coefficient for primary education versus none, in ascending order. Here, in over half of the surveys (36 of 67), the coefficient is in the expected direction (negative). In six surveys—Cameroon 2011, Guyana 2009, Ghana 2008, Benin 2006, Nigeria 2008, and Tanzania 2015-16—the coefficient is both negative and statistically significant. The results in six other surveys show positive and statistically significant coefficients. Thus there is insufficient support for our fourth hypothesis among husbands with primary education.

Figure 3.10 Coefficient for husband having primary education (versus none) from regression on absolute difference between wife's self-reported age and husband's estimate of wife's age

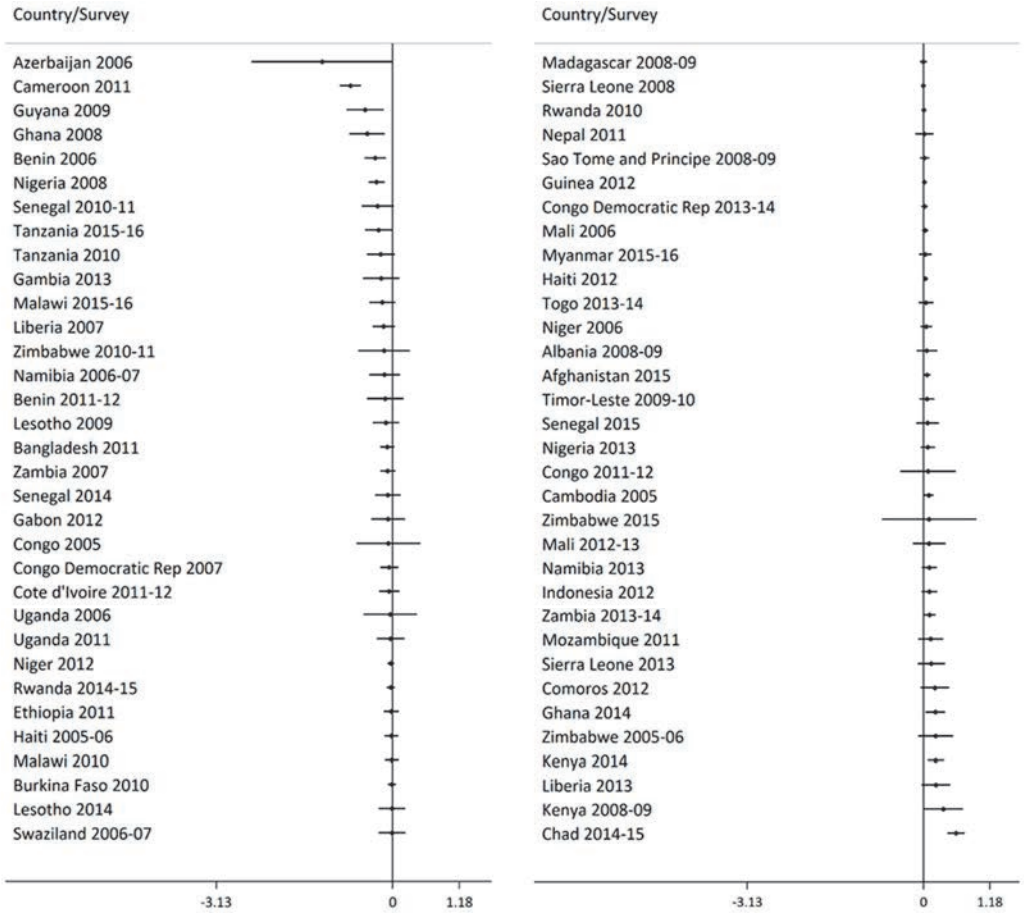
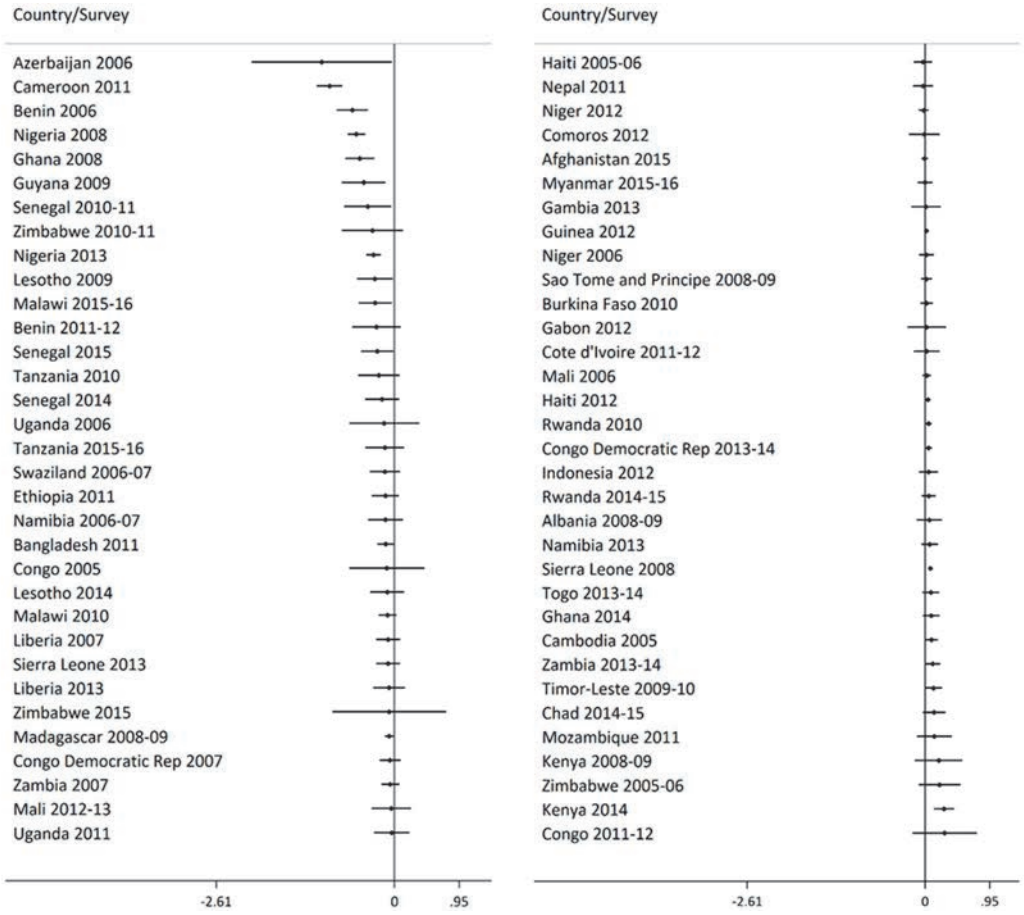


Figure 3.11 displays the effect of husband's secondary-plus education (versus none). As expected, the effect of secondary education on consistency of estimates is stronger than the effect of primary education: it is negative and statistically significant in 13 of 67 surveys, and positive and statistically significant in 6 other surveys, thus providing only limited support for the education-consistency hypothesis among husbands.

Figure 3.11 Coefficient for husband having secondary or higher education (versus none) from regression on absolute difference between wife's self-reported age and husband's estimate of wife's age



4 Conclusion

The goal of this report has been to assess the quality and consistency of age and date reports in DHS surveys conducted since 2000 in 67 countries. It is the most recent of several reports on various aspects of DHS data quality. Ages and dates are well suited to statistical analysis because they are represented numerically and have true and fixed values, even if we do not know exactly what the true values are.

The first chapter described the steps of editing and imputing during fieldwork and data processing. Great care is taken to train and supervise the interviewers to obtain the best possible estimates of ages and dates. The eligibility of adults for the surveys of women and men depends on obtaining accurate values of age near the lower and upper age boundaries within the household survey. The eligibility of young children for the detailed health questions depends on obtaining accurate estimates of when they were born within the surveys of women. Age-specific fertility rates, under-five mortality rates, immunization rates, anthropometry scores, and many other DHS indicators depend on accurate estimates of age.

Appendix A supplements the first chapter with an inventory of all the locations in DHS surveys where the respondents are asked for ages, dates, and durations since an event. This detailed assessment focused on just a few of those locations: the ages of all household members, provided by the household respondent during the household survey; the self-reported ages and birthdates of women and men in the surveys of women and men; women's self-reports of age and date of first union in the survey of women; the birthdates (and ages, if living) of children in the birth histories, provided by the mother; and the women's and men's estimates of their respective spouses' ages in the surveys of women and men. These are a mix of self-reports of one's own age and reports or estimates of someone else's age. It would be possible to analyze the other ages or dates that are asked about, such as age at sterilization, but we focus on ages that we expect to be most salient to the respondents.

The second chapter assessed the ages listed above, other than spousal estimates, in terms of three types of measures: incompleteness, heaping, and transfers. A total of 11 indicators were used. For each indicator, the distribution across all surveys was described and the surveys with the most extreme levels were identified. DHS surveys include explicit incompleteness codes to identify which ages and dates were adjusted during machine editing, and the type of adjustment. There is no incompleteness code for age in the household survey, because only age is asked—not month and year of birth. In the surveys of men and women, age and birthdate are flagged for about 21% of cases, usually because month of birth is not provided or is inconsistent with age, year of birth, and date of interview. Across surveys, the levels of incompleteness for men and for women are very highly correlated (r=0.97). In several surveys the level is considerably lower for men than for women. The reverse pattern is never found, suggesting that knowledge of age and birthdate may be more important for men than for women in some contexts. Women's reports of their age and date at first union have higher levels of incompleteness than their age and birthdate, on average about 31%, mainly because month is missing. Birthdates and ages of all children in the birth histories (not just those born in the past five years) have the lowest levels of incompleteness, on average only 5%. The birth histories involve a good deal of probing and adjusting during fieldwork, often including comparisons with a list of significant national or regional events, and proper sequencing of successive births. Across surveys, incompleteness in the birth histories has little correspondence with incompleteness of women's age and birthdate.

Heaping, measured with Myers' Blended Index, is mostly due to preference at final digits 0 and 5. Heaping is low, only about 5%, on average, in the household surveys (for ages 0-79). Heaping is very conspicuous at ages such as 50, 60, 70, but not at ages below 30 in most surveys. In general, the numbers of cases in the ages with greatest heaping are relatively small, and that is the reason why the heaping index is usually low in the household survey. It tends to be a little larger in the surveys of women and men (for ages 20-49), on

average 6% for women and 7% for men, and only 3% for living children (for ages 0-29) in the birth histories. Ages reported in these other locations are generally the same as ages reported in the household survey; differences in the heaping index are largely due to differences in the age range. Pairwise correlations between heaping indices, across surveys, are high, in a range from 0.79 to 0.91.

Age/date displacement can be measured in different ways, and here we use a measure developed in an earlier assessment (Pullum 2006). The most important potential transfers affect eligibility for further interviewing. It is generally understood that such transfers tend to be made by interviewers, as part of discussion or negotiation with respondents who are not confident in the age of the person they are reporting on. It is estimated that on average only about 3% of girls or women who are actually age 15 are misreported as being age 14, and therefore not eligible for the survey of women. These reports are not made by the 15-year-old herself, because DHS does not interview respondents below age 15, but by a different person who is the household respondent, often the girl's mother. Much more serious, it is estimated that on average 21% of women who are actually age 49 are misreported as being age 50, and therefore not eligible for the survey of women. Some of these transfers are due to heaping at age 50. Displacement of children in the birth histories, to a birthdate that is outside the range for additional questions on child health, is estimated to occur on average for about 8% of the children whose true birthdate is in the 12 months nearest the threshold for those questions. The correspondence among these potential transfers, across surveys, is low.

All of these measures vary substantially. There are many surveys with values close to zero on all measures, and others with very high values. There are some surveys in which month of birth is hardly ever given. Age/date transfers are sometimes large but in a direction opposite to what we would expect, particularly around age 15 or around the date for the health questions, clearly as a result of over-correction during training and supervision. Surveys with extreme values were listed. Summary indices of incompleteness, heaping, and transfers were constructed and tracked over time, indexed by the year of the survey (or the first year of data collection, for surveys spanning two years). The indicators fluctuated substantially from 2000 to 2015 and did not show a systematic trend. A single composite index was constructed for each of the 67 countries. The countries in the highest quintile (with the most problems) and the lowest quintile (with the fewest problems) were identified.

The third chapter investigated the quality and consistency of spousal age estimates compared to self-reports of age, in the surveys that included interviews of men as well as women. Our first set of findings indicated that in the absence of an age gap between spouses, women tend to estimate that their husbands are older than their self-reported age and men tend to estimate that their wives are younger than their self-reported age. Evidence indicates that where there is an age difference between spouses, women tend to estimate in a way that reduces the gap: they underestimate the age of older husbands and overestimate the age of younger husbands. Men underestimate the age of wives who are older than they are, which reduces the gap, but they also tend to underestimate the age of wives who are younger, which increases the gap. The magnitude of underestimation is smaller when wives are younger. The results lend general support to the hypothesis that spousal age estimates tend to reduce the actual age gap between spouses.

The third chapter also examined metrics of age heaping—excess age reports at digits ending in 0 and 5, and Myers' Blended Index (the measure of heaping also used in chapter 2)—for both spousal and self-estimates of age among husbands and wives. We hypothesized that heaping would be greater for spousal reports than for self-reports. Our results gave fairly strong support for this hypothesis: in the vast majority of countries, there was more heaping for estimates of wife's and husband's age than for self-reported age. Our third hypothesis was that the older the individual, the less consistent their spousal age estimate would be with their spouse's reported age. Here the evidence was more mixed. Regression results for women's age on the absolute difference of their husband's age and their report of their husband's age were consistent with this hypothesis in one-fifth of countries and consistent in the reverse scenario in over half of the surveys in which men were asked the age of their wives. The fourth and final hypothesis was that more educated

spouses would provide more consistent estimates of their partner's age than uneducated spouses. The hypothesis was supported modestly among women for their estimates of their husband's age and only very weakly among men.

Each interview is an interaction between an interviewer and a respondent, both of whom contribute to the quality of the age and date responses—the respondent to the degree that she or he actually knows the answer, and the interviewer to the degree that she or he is able to elicit the answer. If the respondent really does not know her or his age, then a skilled interviewer can nevertheless produce a good estimate by referring to historical events on a list or other household members whose age may be better known, but estimates by their nature will tend to be affected by digit preference and/or a desire to comply with guidelines related to heaping and to eligibility boundaries.

Misreporting of ages and dates can potentially affect many of the indicators in DHS surveys. There has not been a thorough inquiry into these effects, but other features of the surveys and indicators mitigate against major distortion. For example, the age range for eligibility for the survey of women is 15 to 49, but at the ends of that range fertility is low and unmet need for contraception is low. The beginning of the time interval for the additional health questions is always at least 60 months before the month of interview, easily long enough to include abundant information about feeding practices, immunizations, child illness and treatment, and so on. Heaping at women's ages ending at 0 and 5 could have some ramifications for age-specific fertility rates, but the standard age intervals (15-19, 20-24, etc.) apply to the mother's age at the child's birth, not to her current age, so the fact that the age intervals begin with a multiple of five is unrelated to any heaping of current age at multiples of five. Errors in reported ages and dates would be more serious if they include bias—for example, if there is a consistent tendency for respondents to be reported as older than they actually are—rather than being heaped by symmetric shifts from both directions. Systematic omission of cases, for example of births that resulted in a neonatal death, will have more serious ramifications than age and date errors.

Problems with the reporting of ages and dates are potentially symptoms of broader data quality issues that are harder to detect. This report does not systematically investigate this assumption, but many of the surveys and countries that have been identified with the highest levels of the misreporting indicators would appear on lists of surveys and countries that have posed the greatest challenges during fieldwork.

It is hoped that future surveys in the countries that have had the most problems in the past can be targeted for special training and supervision of interviewers. There is evidence that displacement and heaping, in particular, can be reduced—but there is also evidence that too much focus on displacement of children or on heaping at final digit 0 can lead to over-correction. The biggest determinant of good age reporting is probably the value, to the respondents, in everyday life, of knowing their ages or the ages of their children. This component of data quality varies from one setting to another and is outside the control of a survey operation.

References

Arnold, F. 1991. *An Assessment of Data Quality in the Demographic and Health Surveys*. Paper read at Demographic and Health Survey Conference, at Washington, DC, USA.

Croft, T.N. 1991. *DHS Data Editing and Imputation*. Calverton, Maryland, USA: Demographic and Health Surveys. Available at https://dhsprogram.com/pubs/pdf/DHSG3/DHS_Data_Editing.pdf.

Curtis, S.L. 1995. *Assessment of the Quality of Data Used for Direct Estimation of Infant and Child Mortality in DHS-II Surveys*. DHS Occasional Papers No. 3. Calverton, Maryland, USA: Macro International. Available at http://dhsprogram.com/pubs/pdf/OP3/OP3.pdf.

Gage, A.J. 1995. *An Assessment of the Quality of Data on Age at First Union, First Birth, and First Sexual Intercourse for Phase II of the Demographic and Health Surveys Programs*. DHS Occasional Papers No. 4. Calverton, Maryland, USA: Macro International. Available at http://dhsprogram.com/pubs/pdf/OP4/OP4.pdf.

ICF. 2017. *Demographic and Health Survey Interviewer's Manual*. Rockville, Maryland, USA: ICF. Available at http://www.dhsprogram.com/pubs/pdf/DHSM1/DHS7-Interviewer's-Manual-EN-12Jun2017-DHSM1.pdf.

ICF International. 2011. *Demographic and Health Surveys Methodology - Questionnaires: Household, Woman's, and Man's*. MEASURE DHS Phase III. Calverton, Maryland, USA: ICF International. Available at http://dhsprogram.com/pubs/pdf/DHSQ6/DHS6_Questionnaires_5Nov2012_DHSQ6.pdf.

ICF International. 2012. *Demographic and Health Survey Interviewer's Manual*. MEASURE DHS Basic Documentation No. 2. Calverton, Maryland, USA: ICF International. Available at http://www.dhsprogram.com/pubs/pdf/DHSM1/DHS6_Interviewer_Manual_19Oct2012_DHSM1.pdf.

ICF International. 2015a. *DHS-7 Biomarker Questionnaire*. DHS Model Questionnaires - Phase 7 (2013-2018). Rockville, Maryland, USA: ICF International. Available at https://dhsprogram.com/pubs/pdf/DHSQ7/DHS7_Biomarker_QRE_EN_09Jun2015_DHSQ7.pdf.

ICF International. 2015b. *DHS-7 Household Questionnaire*. DHS Model Questionnaires - Phase 7 (2013-2018). Rockville, Maryland, USA: ICF International. Available at http://dhsprogram.com/pubs/pdf/DHSQ7/DHS7_Household_QRE_EN_16Mar2017_DHSQ7.pdf.

ICF International. 2015c. *DHS-7 Man's Questionnaire*. DHS Model Questionnaires - Phase 7 (2013-2018). Rockville, Maryland, USA: ICF International. Available at http://dhsprogram.com/pubs/pdf/DHSQ7/DHS7_Mans_QRE_EN_12Oct2015_DHSQ7.pdf.

ICF International. 2015d. *DHS-7 Woman's Questionnaire*. DHS Model Questionnaires - Phase 7 (2013-2018). Rockville, Maryland, USA: ICF International. Available at http://dhsprogram.com/pubs/pdf/DHSQ7/DHS7-Womans-QRE-EN-07Jun2017-DHSQ7.pdf.

Johnson, K., M. Grant, S. Khan, Z. Moore, A. Armstrong, and Z. Sa. 2009. *Fieldwork-Related Factors and Data Quality in the Demographic and Health Surveys Program*. DHS Analytical Studies No. 19. Calverton, Maryland, USA: ICF Macro. Available at http://dhsprogram.com/pubs/pdf/AS19/AS19.pdf.

Macro International. 1995a. *DHS Model "A" Questionnaire with Commentary for High Contraceptive Prevalence Countries*. Calverton, Maryland, USA: Macro International. Available at https://dhsprogram.com/pubs/pdf/DHSQ3/DHS-III-Model-A.pdf.pdf.

Macro International. 1995b. *DHS Model "B" Questionnaire with Commentary for Low Contraceptive Prevalence Countries*. Calverton, Maryland, USA: Macro International. Available at https://dhsprogram.com/pubs/pdf/DHSQ3/DHS-III-Model-B.pdf.pdf.

Macro International. 2008a. *Household Questionnaire*. DHS Model Questionnaires - Phase 5 (2003-2008). Calverton, Maryland, USA: Macro International. Available at http://dhsprogram.com/pubs/pdf/DHSQ5/DHS5-Household-QRE-15-Oct-2008.pdf.

Macro International. 2008b. *Man's Questionnaire*. DHS Model Questionnaires - Phase 5 (2003-2008). Calverton, Maryland, USA: Macro International. Available at http://dhsprogram.com/pubs/pdf/DHSQ5/DHS5-Man's-QRE-22-Aug-2008.pdf.

Macro International. 2008c. *Woman's Questionnaire*. DHS Model Questionnaires - Phase 5 (2003-2008). Calverton, Maryland, USA: Macro International. Available at http://dhsprogram.com/pubs/pdf/DHSQ5/DHS5-Woman's-QRE-22-Aug-2008.pdf.

MEASURE DHS, and ICF International. 2013. *Standard Recode Manual for DHS-6*. Calverton, Maryland, USA: ICF International.

ORC Macro. 2001a. *Model "A" Questionnaire with Commentary for High Contraceptive Prevalence Countries*. ORC Macro. Available at http://dhsprogram.com/pubs/pdf/DHSQ4/DHS-IV-Model-A.pdf.pdf.

ORC Macro. 2001b. *Model "B" Questionnaire with Commentary for Low Contraceptive Prevalence Countries*. ORC Macro. Available at http://dhsprogram.com/pubs/pdf/DHSQ4/DHS-IV-Model-B.pdf.pdf.

Pullum, T.W. 2006. *An Assessment of Age and Date Reporting in the DHS Surveys, 1985-2003*. DHS Methodological Reports No. 5. Calverton, Maryland, USA: Macro International. Available at http://dhsprogram.com/pubs/pdf/MR5/MR5.pdf.

Pullum, T.W., and S. Becker. 2014. *Evidence of Omission and Displacement in DHS Birth Histories*. DHS Methodological Reports No. 11. Rockville, Maryland, USA: ICF International. Available at http://dhsprogram.com/pubs/pdf/MR11/MR11.pdf.

Rutstein, S.O., G.T. Bicego, A.K. Blanc, N. Rutenberg, F. Arnold, and J.A. Sullivan. 1990. *An Assessment of DHS-I Data Quality*. DHS Methodological Reports No. 1. Columbia, Maryland, USA: Institute for Resource Development/Macro Systems Inc. Available at http://dhsprogram.com/pubs/pdf/MR1/MR1.pdf.

Schoumaker, B. 2014. *Quality and Consistency of DHS Fertility Estimates, 1990 to 2012*. DHS Methodological Reports No. 12. Rockville, Maryland, USA: ICF International. Available at http://dhsprogram.com/pubs/pdf/MR12/MR12.pdf.

Appendix A All Ages, Dates, and Durations Asked of DHS Respondents and their Evolution between DHS Phases 4 to 7

During each five-year phase of the project, The DHS Program has published model household, women's, and men's questionnaires that form the basis of standard data collection during that phase. Known as 'core questionnaires', these are adapted for each country during each survey phase. Hence, questions for individual surveys—located in the back of each final survey report—may vary slightly from what is described here. However, collection of most basic indicators including ages and dates is almost fully standardized throughout each five-year phase.

In the period covered by this report—from 2000 until present—The DHS Program has spanned four phases: DHS-4 (1997-2003), DHS-5 (2003-2008), DHS-6 (2008-2013), and DHS-7 (2013-present). The collection of information related to ages and dates in each questionnaire type during all four phases is described in the next three sections of this appendix. Note that in DHS-4 there were two types of questionnaires: Model "A" for countries with high contraceptive prevalence (ORC Macro 2001a) and Model "B" for countries with low contraceptive prevalence (ORC Macro 2001b); in subsequent program phases, questionnaires for high and low contraceptive prevalence countries were unified into single questionnaires for households, women, and men.

Household and Biomarker Questionnaires

In DHS-4, the core household questionnaire for both Model "A" and Model "B" countries asks an adult household representative to list the name (Q2) and age (Q7) of all household members, to identify a head of household, and to describe each member's relationship to the household head (Q3). Additionally, the weight and height measurement section, which is located at the end of the household questionnaire, asks for the date of birth of each household child under age 6 (Q39). However this section is not necessarily intended to be filled in during the household interview, as the measurements require separate consent from each woman and are typically done by a biomarker specialist.

In DHS-5 to DHS-7, the model household questionnaires continue to ask the household representative for the names and ages of all household members and their relationship to the head of household (ICF International 2011, 2015b; Macro International 2008a). Starting in DHS-5, the household questionnaire section on height, weight, and hemoglobin instructs interviewers that, if the mother of children age 0 to 5 was interviewed, they should copy the dates of birth from the mother's questionnaire (Q503), otherwise they should ask an adult respondent for the child's birth date. This practice continued in DHS-6 and DHS-7. In DHS-7 weight, height, and hemoglobin measurement were moved to a separate biomarker questionnaire (ICF International 2015a).

Woman's Questionnaire

In the DHS-4 Model "A" and Model "B" women's questionnaires (ORC Macro 2001a, 2001b), interviewed women are asked the following items related to ages and dates:

- Duration in years of continuous residence at current city, town, or village (Q103)
- Month and year of birth (Q105)
- Age at last birthday (Q106)
- Month and year of birth for each child ever born (Q215)

- For all living children, the age of each child at the child's last birthday (Q217)
- For all children who have died, the age of each child at death in days, months, or years (Q220)
- If currently pregnant, the number of completed months pregnant (Q227)
- If have ever had a pregnancy that did not end in a live birth, the month and year the last such pregnancy ended (Q230); if ended in past five years[1], the number of months that pregnancy lasted (Q232)
- For any and all other pregnancies in the past five years that did not result in a live birth, the month and year the pregnancy ended and the number of months each pregnancy lasted (Q234)
- If more than five years before the survey there were any other pregnancies that did not result in a live birth, the month and year that the last such pregnancy ended (Q236)
- The number of days, weeks, months, or years since the respondent's last menstrual period started, or the date it started (Q237)
- If sterilized, month and year when sterilized (Q316)
- If currently using a method, month and year when started using method without stopping (Q316A)
- Contraceptive calendar: starting month and year and ending month and year of each period of contraceptive use or non-use for the past five years (Model "A" only: Q317 and calendar page)
- If have given birth in the past five years and received any antenatal care, number of months pregnant when first received antenatal care for the most recent pregnancy (Q408); if received antenatal care more than once during that pregnancy, number of months pregnant when last received antenatal care (Q411)
- If have given birth in the past five years and a health professional or traditional birth attendant checked on respondent's health after the most recent child was born, the number of days or weeks after delivery that the first check took place (Q430)
- If have given birth in the past five years and period has since resumed, the number of months after most recent birth that the respondent did not have a period (Q436, column 1); if have given birth more than once in the past five years and period resumed between second-to-last birth and most recent pregnancy, the number of months after second-to-last birth that the respondent did not have a period (Q436, column 2)
- If have resumed sexual relations since birth of child in the past five years, number of months abstinent after birth (Q439, column 1); if have given birth more than once in past five years, number of months after second-to-last birth the respondent abstained from sex (Q439, column 2)
- If have breastfed either of last two children born in past five years, number of hours or days after birth first put child to breast (Q441, columns 1 and 2)
- If not still breastfeeding last two children born in past five years, number of months each child was breastfed (Q446, columns 1 and 2)

[1] Here and elsewhere, 'past five years' means that the questionnaire refers to the period starting January of the calendar year five years prior to the year of fieldwork; depending on the dates of fieldwork this may span up to six years before the date of interview. Consult individual survey reports for details on dates of fieldwork and the dates of reference periods used in the country-specific questionnaires.

- If ever married or lived with man as if married, month and year started living with first husband/partner (Q511); if the year first started living with first/husband partner is unknown, age when started living with first husband/partner (Q512)
- Marital calendar: month and year of beginning and end of each union for the past five years (Model "A" only: Q513 and calendar page)
- If ever had sex, age at first sex (Q514)
- If ever had sex, time since last sex in days, weeks, months, or years (Q515)
- If last sex was with a man who is not a husband or cohabiting partner, duration of sexual relationship in days, weeks, months, or years (Q518)
- If have had sex in past 12 months with any other man who is not a husband or cohabiting partner, for the next-most-recent sexual partner, the duration of that sexual relationship in days, months, weeks, or years (Q522)
- If currently married or living with a man as if married, husband/partner's age at last birthday (Q702)

Dates of immunization for each child age 5 and under are recorded from the vaccination card. If the card is not available, the dates of immunization(s) are not asked. DHS interviewers are instructed to reconcile a woman's reported age and date of birth (see Figure 1.1 "Field imputation procedures from The DHS Program's interviewer training manual" for more details). However the questionnaire does not include any formal instructions to reconcile disparities between child's age and date of birth.

The DHS-5 woman's core questionnaire (Macro International 2008c) includes all DHS-4 age and date questions at some point in the questionnaire, with the exception of the marital calendar, the number of months pregnant when last received antenatal care, and the time after birth first breastfed second-to-last baby born in past five years. Additionally, women are asked a number of additional timing questions related to antenatal care and post-delivery care for the most recent birth in the past five years. They are asked the month and year of last tetanus injection (if any) prior to most recent pregnancy that resulted in a birth in the past five years (Q419); if year is unknown, they are asked how many years ago they received the injection (Q420). If, for the most recent birth in the past five years the woman delivered at a health facility, she is asked the number of hours, days, or weeks she stayed there after birth (Q437); if someone at the facility checked on her health prior to discharge she is asked the number of hours, days, or weeks after birth that the first check took place (Q440). The original question about how long after birth a health care provider checked on the woman's health (Q430 in DHS-4 woman's questionnaire) is asked slightly differently for two groups of women: women who gave birth in a facility are asked about the time after delivery that a health care provider or traditional birth attendant first checked on their health post-discharge, while women who did not deliver in a facility are asked about the time after delivery that a health care provider or traditional birth attendant first checked on their health, in hours, days, or weeks (Q445 for both groups). If, for the most recent birth in the past five years, a health care provider or traditional birth attendant checked on the child's health within the first two months after birth, women are asked how many hours, days, or weeks after birth the first check took place (Q450). Also, in DHS-5, questions about resumption of menstrual period after birth and duration of abstinence after birth are asked for up to three births in the past five years (Q456, Q459, columns 1 to 3).

Additional questions about dates and timing asked in the DHS-5 woman's questionnaire relate to sexual history and HIV testing. In DHS-5, women are asked not only the time since last sexual intercourse but also the time since last sexual intercourse, for up to two most recent additional sexual partners in the past 12 months (Q627 columns 2 and 3), the duration of that sexual relationship (Q631, columns 1 to 3), and ages of the three most recent sexual partners in the past 12 months (Q633, columns 1 to 3). If age is unknown, the respondent is asked whether the partner is older, younger, or the same age (Q624) and, if older, whether

there is at least a 10-year age gap (Q625). If the respondent has given birth in the past five years and was tested for HIV[2] during antenatal care for that pregnancy but has been tested since then, she is asked whether the last time she was tested is less than 12 months ago, 12 to 23 months, or 2 years or more (Q921). If the respondent was not tested during her pregnancy or has not had a birth in the past five years but has ever been tested for HIV, she is asked whether the last time she was tested for HIV is less than 12 months ago, 12 to 23 months, or 2 years or more (Q923).

In the DHS-6 woman's core questionnaire (ICF International 2011), all age and date questions from DHS-5 are asked, with the exception of the duration at current residence and the month and year of the last tetanus injection prior to pregnancy that resulted in a birth in the past five years. Instead, women who have given birth in the past five years are asked the number of years before their pregnancy they last received a tetanus injection. Additionally, the separate questions about health checkups for the respondent after most recent delivery at the facility and after discharge are combined—as in DHS-4—into a single question for all women who have given birth in the past five years about how long after delivery a health professional or traditional birth attendant checked on her health (Q440). For the last three sexual partners in the past 12 months, instead of the duration of the sexual relationship, women are asked how many days, weeks, months, or years ago they first had sexual intercourse with this person (Q622). If age of the last three sexual partners in the past 12 months is unknown, women are no longer asked to estimate whether the partner is older, younger, or the same age. For HIV testing within the past two years, women are asked to specify the number of months since the most recent test (Q925 and Q927).

In the DHS-7 woman's core questionnaire (ICF International 2015d), women are asked all age and date questions from the DHS-6 core questionnaire except duration of breastfeeding. Additionally, they are asked the duration in years of continuous residence in current city, town, or village (Q102), and the day of birth for all children ever born (Q215). For respondents whose most recent birth in the past five years was at a health facility, the question about health checkups after the most recent delivery for the respondent is divided into two questions—as in DHS-5—one about the amount of time after delivery someone checked on her health inside the facility if at all (Q436), and another about the amount of time after delivery that someone checked on her health post-discharge if at all (Q442). As in all other survey rounds described here, women who gave birth in the past five years but delivered their most recent birth outside of a health facility are asked a single question about the amount of time after delivery their first health check took place if at all (Q450). New in DHS-7, respondents who have most recently given birth in the past five years at a health facility are asked two separate questions about health checks on the baby in the first two months after birth: one about the amount of time after birth someone checked on the baby's health while at the facility if at all (Q439), and another about the amount of time after birth that someone checked on the baby's health post-discharge if at all (Q446). Women who whose most recent birth in the past five years was outside of a health facility are asked a separate question about the timing of the first health checkup for the baby if it occurred within the first two months after birth (Q454). In DHS-7—as in DHS-4—questions about number of months after birth the respondent did not have her period (Q460), and did not have sexual intercourse (Q463), are only asked for the two most recent live births in the past five years.

Man's Questionnaire

DHS surveys have included men as early as 1987, however the DHS core questionnaires did not include a model man's questionnaire until DHS-5, in 2003. In the DHS-5 model man's questionnaire (Macro International 2008b), men are asked for the following information related to age and dates:

[2] In DHS-5, referred to as 'tested to see if you have the AIDS virus.' In DHS-6, the preface question is 'tested to see if you have the AIDS virus' and the subsequent question is 'How many months ago was your most recent HIV test.' In DHS-7, referred to as HIV throughout.

- Duration of continuous residence in current home (Q102)
- Month and year of birth (Q106)
- Age at last birthday (Q107)
- If have had children, age when first child was born (Q212)
- If have any living children, age of youngest child (Q214)
- Age of all current wives/partners at last birthday (Q408)
- If ever married or lived with woman as if married, month and year started living with first wife/partner (Q411); if year began living with first wife/partner is unknown, age began living with first wife/partner (Q412)
- If ever had sex, age at first sexual intercourse (Q414)
- If ever had sex, number of days, weeks, months, or years since last sexual intercourse (Q419); if have had sex with more than one partner in the past 12 months, the number of days, weeks, or months since last sexual intercourse with the next two most recent partners (Q421, columns 2 and 3)
- If any of the last three sexual partners in the past 12 months is not a wife or live-in partner, the duration of the sexual relationship with that partner in days, months, or years (Q425, columns 1 to 3)
- If ever tested for HIV[3], whether the time since last test was less than 12 months ago, 12 to 23 months ago, or 2 or more years ago (Q714)

For the model man's questionnaire, DHS interviewers are instructed to reconcile any conflicts between men's reported age and date of birth, as with the model woman's questionnaire.

In the DHS-6 man's core questionnaire, men are not asked their duration of current residence, but they are asked the age of their last three sexual partners in the past 12 months (Q425, columns 1 to 3). Instead of the duration of the sexual relationship with the last three non-marital partners in the past 12 months, they are asked how long ago they first had sexual intercourse with each partner, regardless of whether married to them or not (Q423, columns 1 to 3). And, instead of categorical values for time since last HIV test, if the last test was less than two years ago men are asked the number of months since their last HIV test (Q713). Additionally, if men are circumcised, they are asked their age at circumcision (Q802).

In the DHS-7 man's questionnaire (ICF International 2015c), men are asked all of the age and date questions asked in DHS-6. Additionally, they are asked the duration in years of continuous residence in current city, town, or village (Q102).

[3] See previous note about use of the term 'AIDS virus' in earlier survey rounds.

Appendix B Example National Historical Events Calendar Used for Date Estimation during the 2012-13 Pakistan DHS

Main Events in Pakistan

Date	Event	Date	Event
Oct. 27, 1958	Field Marshal Ayub Khan imposed **Martial Law** in Pakistan	Nov. 13, 1993	Farooq Leghari elected **President**
Sep. 6, 1965	**War** between Pakistan & India	Nov. 5, 1996	Benazir Bhutto's government sacked
Mar. 26, 1969	General Yahya Khan imposed **Martial Law**	Feb. 17, 1997	Nawaz Sharif became **Prime Minister** (Second time)
Dec. 16, 1971	**War** between Pakistan & India (Fall of Dhaka)	May 28, 1998	Pakistan made first nuclear test
Dec. 20, 1971	Mr. Zulfikar Ali Bhutto became President of Pakistan	May/Jun. 1999	Kargil event
Aug./Sep. 1973	Floods in Punjab	Oct. 12, 1999	Gen. Pervez Musharraf impose **Martial Law**
Jun./Jul. 74 & 76	Floods in Sindh	Aug. 2000	Local Bodies Elections (**Nazims** system)
Jul. 5, 1977	Gen. Zia-ul-Haq declared **Martial Law**	Sep. 11, 2001	Attack on **World Trade Centre** (USA)
Apr. 4, 1979	Z. A. Bhutto was **hanged**	Oct. 2002	Provincial & National Assemblies Elections
1985	Mr. Muhammad Khan Jonaijo became Prime Minister	Oct. 2002	Attacked on **Afghanistan** by USA
Dec. 1985	Sohrab Goth clean-up operation in Karachi	Mar. 18, 2003	Attacked on **Iraq** by USA
Apr. 10, 1988	Disaster of **Ojri Camp** in Rawalpindi	Sep. 2005	2nd Local Bodies Elections (**Nazims** system)
Aug. 17, 1988	Gen. Zia's **plane crash**	Oct, 2005	**Earthquake** in AJK and parts of Pakistan
Dec. 2, 1988	Benazir Bhutto became **Prime Minister**.(First Time)	Dec. 27, 2007	Banazir Bhutto was **Murdered** in Rawalpindi
Dec. 13, 1988	Ghulam Ishaq Khan became **President**	Feb. 2008	**Election** of National and Provincial Assemblies
Aug. 6, 1990	Benazir Bhutto's government sacked	Mar. 2008	**Syed Yousaf Raza Gillani** elected as Prime Minister
Nov. 6, 1990	Nawaz Sharif elected **Prime Minister** (First time)	2008	**Asif Ali Zardari** elected as President of Pakistan
Mar. 22-23, 1992	Pakistan won the final of **Cricket World Cup**	Jul./Aug. 2010	Floods in Pakistan
Aug./Sep. 1992	Floods in Punjab and Sindh	2012	Syed Yousaf Raza Gillani's Government **dismissed by the Supreme Court**
Apr. 18, 1993	Nawaz Sharif's government dismissed	2012	**Raja Parvez Ashraf** elected as Prime Minister
Oct. 19, 1993	Benazir Bhutto became **Prime Minister** (Second time)		

Appendix C List of Surveys Analyzed in Chapter 2

Survey	Household	Women	Men	Children
Afghanistan 2015	203,708	29,461	10,760	114,905
Albania 2008-09	31,099	7,584	3,013	12,208
Angola 2015-16	74,902	14,379	5,684	37,223
Armenia 2000	26,371	6,430	1,719	10,517
Armenia 2005	25,028	6,566	1,447	9,734
Armenia 2010	23,629	5,922	1,584	8,067
Azerbaijan 2006	30,637	8,444	2,558	12,465
Bangladesh 2004	55,883	11,440	4,297	28,285
Bangladesh 2007	53,413	10,996	3,771	26,104
Bangladesh 2011	83,731	17,842	3,997	40,253
Bangladesh 2014	81,624	17,863	-	38,894
Benin 2001	30,417	6,219	2,709	15,689
Benin 2006	90,650	17,794	5,321	47,808
Benin 2011-12	88,174	16,599	5,180	42,617
Bolivia 2003	81,669	17,654	6,230	39,272
Bolivia 2008	77,757	16,939	6,054	35,834
Burkina Faso 2003	60,766	12,477	3,605	33,006
Burkina Faso 2010	82,095	17,087	7,307	46,357
Burundi 2010	42,420	9,389	4,280	20,682
Cambodia 2000	66,285	15,351	-	34,696
Cambodia 2005	73,010	16,823	6,731	34,854
Cambodia 2010	76,920	18,754	8,239	33,437
Cambodia 2014	74,122	17,578	5,190	30,443
Cameroon 2004	51,976	10,656	5,280	24,747
Cameroon 2011	72,622	15,426	7,191	36,057
Chad 2004	29,614	6,085	1,887	16,914
Chad 2014-15	99,620	17,719	5,248	57,995
Colombia 2000	47,520	11,585	-	20,117
Colombia 2005	157,840	41,344	-	67,390
Colombia 2010	204,459	53,521	-	86,947
Colombia 2015	162,459	38,718	35,783	59,761
Comoros 2012	24,499	5,329	2,167	10,700
Congo 2005	31,481	7,051	3,146	14,519
Congo Democratic Republic 2007	48,291	9,995	4,757	24,255
Congo 2011-12	51,449	10,819	5,145	28,337
Congo Democratic Republic 2013-14	95,949	18,827	8,656	50,637
Cote d'Ivoire 2011-12	51,187	10,060	5,135	23,870
Dominican Republic 2002	110,758	23,384	2,833	49,566
Dominican Republic 2007	123,738	27,195	27,975	54,136
Dominican Republic 2013	41,267	9,372	10,306	16,942
Egypt 2000	91,173	15,573	-	47,982
Egypt 2005	112,710	19,474	-	55,411
Egypt 2008	92,120	16,527	-	45,196
Egypt 2014	120,276	21,762	-	56,155
Ethiopia 2000	68,642	15,367	2,607	33,741
Ethiopia 2005	67,540	14,070	6,033	32,591
Ethiopia 2011	77,744	16,515	14,110	37,796
Gabon 2000	32,391	6,183	2,004	14,934
Gabon 2012	41,675	8,422	5,654	20,903
Gambia 2013	52,691	10,233	3,821	24,163
Ghana 2003	26,601	5,691	5,015	12,976
Ghana 2008	46,536	4,916	4,568	10,378
Ghana 2014	43,945	9,396	4,388	20,678

Survey	Household	Women	Men	Children
Guatemala 2014-15	102,510	25,914	11,145	51,876
Guinea 2005	38,182	7,954	3,174	20,876
Guinea 2012	45,049	9,142	3,782	22,957
Guyana 2009	22,845	4,996	3,522	10,197
Haiti 2000	47,361	10,159	3,171	21,963
Haiti 2005-06	47,319	10,757	4,958	21,481
Haiti 2012	59,746	14,287	9,493	25,468
Honduras 2005-06	93,867	19,948	-	46,725
Honduras 2011-12	100,555	22,757	7,120	46,456
India 2005-06	534,161	124,385	74,369	229,017
Indonesia 2002-03	149,222	29,483	8,310	71,607
Indonesia 2007	178,843	32,895	8,758	76,392
Indonesia 2012	185,345	45,607	9,306	76,043
Jordan 2002	46,755	6,006	-	24,051
Jordan 2007	82,470	10,876	-	41,828
Jordan 2012	80,822	11,352	-	40,705
Kenya 2003	37,612	8,195	3,578	19,319
Kenya 2008-09	38,515	8,444	3,465	20,137
Kenya 2014	153,840	31,079	12,819	76,876
Kyrgyz Republic 2012	35,805	8,208	2,413	15,445
Lesotho 2004	40,490	7,095	2,797	13,022
Lesotho 2009	44,546	7,624	3,317	12,903
Lesotho 2014	40,197	6,621	2,931	10,540
Liberia 2007	34,670	7,092	6,009	17,826
Liberia 2013	48,219	9,239	4,118	25,016
Madagascar 2003-04	38,325	7,949	2,432	18,249
Madagascar 2008-09	85,858	17,375	8,586	42,922
Malawi 2000	63,823	13,220	3,092	31,402
Malawi 2004	60,747	11,698	3,261	29,156
Malawi 2010	118,850	23,020	7,175	60,461
Malawi 2015-16	120,492	24,562	7,478	60,399
Maldives 2009	42,050	7,131	1,727	18,598
Mali 2001	66,505	12,849	3,405	35,778
Mali 2006	73,685	14,583	4,207	39,972
Mali 2012-13	58,330	10,424	4,399	29,585
Morocco 2003-04	64,044	16,798	-	29,098
Mozambique 2003	63,496	12,418	2,900	29,419
Mozambique 2011	62,750	13,745	4,035	32,312
Myanmar 2015-16	55,584	12,885	4,737	20,428
Namibia 2000	31,675	6,755	2,954	13,748
Namibia 2006-07	42,633	9,804	3,915	17,836
Namibia 2013	41,646	10,018	4,481	16,847
Nepal 2001	47,523	8,726	2,261	24,279
Nepal 2006	44,057	10,793	4,397	22,891
Nepal 2011	49,791	12,674	4,121	23,786
Nicaragua 2001	61,351	13,060	-	31,206
Niger 2006	47,964	9,223	3,549	26,853
Niger 2012	64,011	11,160	3,928	35,983
Nigeria 2003	35,820	7,620	2,346	17,736
Nigeria 2008	156,809	33,385	15,486	83,640
Nigeria 2013	178,894	38,948	17,359	97,736
Pakistan 2006-07	727,493	10,023	-	34,759
Pakistan 2012-13	94,169	13,558	3,134	44,832
Peru 2000	131,062	27,843	-	58,396
Peru 2004-08	187,286	41,648	-	81,741
Peru 2009	105,225	24,212	-	46,451

Survey	Household	Women	Men	Children
Peru 2010	101,409	22,947	-	43,404
Peru 2011	98,662	22,517	-	42,890
Peru 2012	103,211	23,888	-	44,139
Philippines 2003	61,864	13,633	4,766	28,430
Philippines 2008	60,901	13,594	-	26,851
Philippines 2013	71,893	16,155	-	30,121
Moldova 2005	32,110	7,440	2,508	9,483
Rwanda 2000	45,247	10,421	2,717	21,612
Rwanda 2005	47,851	11,321	4,820	24,042
Rwanda 2010	56,505	13,671	6,329	27,682
Rwanda 2014-15	54,905	13,497	6,217	26,391
Sao Tome and Principe 2008-09	13,430	2,615	2,296	6,852
Senegal 2005	69,059	14,602	3,761	33,494
Senegal 2010-11	77,269	15,688	4,929	36,828
Senegal 2012-13	41,593	8,636	-	19,977
Senegal 2014	40,723	8,488	3,371	19,964
Senegal 2015	41,902	8,851	3,734	20,727
Sierra Leone 2008	41,985	7,374	3,280	17,003
Sierra Leone 2013	75,299	16,658	7,262	37,294
Swaziland 2006-07	22,143	4,987	4,156	10,113
Tajikistan 2012	38,805	9,656	-	18,598
Timor-Leste 2009-10	67,834	13,137	4,076	31,922
Togo 2013-14	46,577	9,480	4,476	22,914
Turkey 2003	47,894	8,075	-	20,413
Uganda 2000-01	37,951	7,246	1,962	19,444
Uganda 2006	45,439	8,531	2,503	24,822
Uganda 2011	44,977	8,674	2,295	24,618
Ukraine 2007	34,123	6,841	3,178	7,762
Tanzania 2004-05	49,921	10,329	2,635	25,750
Tanzania 2010	50,414	10,139	2,527	26,011
Tanzania 2015-16	64,880	13,266	3,514	33,234
Vietnam 2002	31,529	5,665	-	13,538
Yemen 2013	120,923	25,434	-	58,905
Zambia 2001-02	38,089	7,658	2,145	19,427
Zambia 2007	35,562	7,146	6,500	18,032
Zambia 2013-14	83,058	16,411	14,773	43,234
Zimbabwe 2005-06	42,698	8,907	7,175	17,834
Zimbabwe 2010-11	41,946	9,171	7,480	17,674
Zimbabwe 2015	43,706	9,955	8,396	19,054

Appendix D List of Surveys Analyzed in Chapter 3

Survey	Women asked to state their husband's age	Men asked to state their wife's age	Men's age range		Women's age range	
			For heaping analysis	For all other analysis	For heaping analysis	For all other analysis
Afghanistan 2015	•	•	20-49	15-49	20-49	15-49
Albania 2008-09	•	•	21-49	21-49	20-49	16-49
Armenia 2000	•		25-54	19-54	-	15-49
Armenia 2005	•		20-49	20-49	-	16-49
Armenia 2010	•		20-49	19-49	-	18-49
Azerbaijan 2006	•	•	20-59	18-59	20-49	16-49
Bangladesh 2004	•		25-54	16-54	-	13-49
Bangladesh 2007	•		25-54	18-54	-	15-49
Bangladesh 2011	•	•	25-54	16-54	20-49	13-49
Benin 2001	•		25-64	16-64	-	15-49
Benin 2006	•	•	25-64	17-64	20-49	15-49
Benin 2011-12	•	•	25-64	16-64	20-49	15-49
Bolivia 2003	•		25-63	17-63	-	15-49
Bolivia 2008	•		25-64	15-64	-	15-49
Burkina Faso 2003	•		21-59	19-59	-	15-49
Burkina Faso 2010	•	•	20-59	18-59	20-49	15-49
Burundi 2010	•	•	20-59	15-59	20-49	16-49
Cambodia 2005	•	•	20-49	15-49	20-49	15-49
Cambodia 2014	•		20-49	17-49	-	15-49
Cameroon 2004	•		21-59	18-59	-	15-49
Cameroon 2011	•	•	21-59	17-59	20-49	15-49
Chad 2004	•		20-59	18-59	-	15-49
Chad 2014-15	•	•	20-59	17-59	20-49	15-49
Comoros 2012	•	•	20-59	19-59	20-49	15-49
Congo 2005	•	•	20-59	16-59	20-49	15-49
Congo 2011-12	•	•	20-59	17-59	20-49	15-49
Congo Democratic Republic 2007	•	•	20-59	16-59	20-49	15-49
Congo Democratic Republic 2013-14	•	•	20-59	16-59	20-49	15-49
Cote d'Ivoire 2011-12	•	•	20-59	18-59	20-49	15-49
Dominican Republic 2002	•		20-59	16-59	-	15-49
Dominican Republic 2007	•		20-59	15-59	-	15-49
Dominican Republic 2013	•		20-59	16-59	-	15-49
Ethiopia 2000	•		20-59	18-59	-	15-49
Ethiopia 2005	•		20-59	16-59	-	15-49
Ethiopia 2011	•	•	20-59	15-59	20-49	15-49
Gabon 2000	•		20-59	15-59	-	15-49
Gabon 2012	•	•	20-59	17-59	20-49	15-49
Gambia 2013	•	•	20-59	20-59	20-49	15-49
Ghana 2003	•		20-59	18-59	-	15-49
Ghana 2008	•	•	20-59	18-59	20-49	15-49
Ghana 2014	•	•	20-59	19-59	20-49	15-49
Guinea 2005	•		21-59	18-59	-	15-49
Guinea 2012	•	•	22-59	17-59	20-49	15-49
Guyana 2009	•	•	20-49	17-49	20-49	15-49
Haiti 2000	•		21-59	18-59	-	15-49
Haiti 2005-06	•	•	20-59	17-59	20-49	16-49
Haiti 2012	•	•	20-59	17-59	20-49	15-49
India 2005-06	•		25-54	15-54	-	15-49
Indonesia 2002-03	•		25-54	16-54	-	15-49
Indonesia 2007	•		25-54	17-54	-	15-49

Survey	Women asked to state their husband's age	Men asked to state their wife's age	Men's age range		Women's age range	
			For heaping analysis	For all other analysis	For heaping analysis	For all other analysis
Indonesia 2012	•	•	25-54	16-54	20-49	15-49
Kenya 2003	•		25-54	18-54	-	15-49
Kenya 2008-09	•	•	25-54	17-54	20-49	15-49
Kenya 2014	•	•	25-54	17-54	20-49	15-49
Kyrgyz Republic 2012	•		20-49	19-49	-	16-49
Lesotho 2004	•		21-59	18-59	-	15-49
Lesotho 2009	•	•	21-59	16-59	20-49	15-49
Lesotho 2014	•	•	20-59	17-59	20-49	15-49
Liberia 2007	•	•	20-49	18-49	20-49	15-49
Liberia 2013	•	•	20-49	18-49	20-49	15-49
Madagascar 2003-04	•		20-59	16-59	-	15-49
Madagascar 2008-09	•	•	20-59	15-59	20-49	15-49
Malawi 2000	•		25-54	17-54	-	15-49
Malawi 2004	•		25-54	17-54	-	15-49
Malawi 2010	•	•	25-54	16-54	20-49	15-49
Malawi 2015-16	•	•	25-54	17-54	20-49	15-49
Maldives 2009	•		25-64	19-64	-	18-49
Mali 2001	•		21-59	19-59	-	15-49
Mali 2006	•	•	21-59	20-59	20-49	15-49
Mali 2012-13	•	•	21-59	20-59	20-49	15-49
Moldova 2005	•		20-59	18-59	-	15-49
Mozambique 2003	•		25-64	18-64	-	15-49
Mozambique 2011	•	•	25-64	15-64	20-49	15-49
Myanmar 2015-16	•	•	20-49	16-49	20-49	15-49
Namibia 2000	•		20-59	17-59	-	15-49
Namibia 2006-07	•	•	20-49	17-49	20-49	15-49
Namibia 2013	•	•	25-64	18-64	20-49	15-64
Nepal 2001	•		20-59	15-59	-	15-49
Nepal 2006	•		20-59	15-59	-	15-49
Nepal 2011	•	•	20-49	17-49	20-49	15-49
Niger 2006	•	•	22-59	18-59	20-49	15-49
Niger 2012	•	•	20-59	16-59	20-49	15-49
Nigeria 2003	•		22-59	18-59	-	15-49
Nigeria 2008	•	•	20-59	17-59	20-49	15-49
Nigeria 2013	•	•	20-49	16-49	20-49	15-49
Philippines 2003	•		25-54	17-54	-	15-49
Rwanda 2000	•		20-59	18-59	-	16-49
Rwanda 2005	•		20-59	18-59	-	17-49
Rwanda 2010	•	•	20-59	18-59	20-49	18-49
Rwanda 2014-15	•	•	20-59	18-59	20-49	16-49
Sao Tome and Principe 2008-09	•	•	20-59	18-59	20-49	15-49
Senegal 2005	•		20-59	19-59	-	15-49
Senegal 2010-11	•	•	20-59	17-59	20-49	15-49
Senegal 2014	•	•	20-59	18-59	20-49	15-49
Senegal 2015	•	•	21-59	19-59	20-49	15-49
Sierra Leone 2008	•	•	20-59	18-59	20-49	15-49
Sierra Leone 2013	•	•	20-59	16-59	20-49	15-49
Swaziland 2006-07	•	•	20-49	19-49	20-49	15-49
Tanzania 2004-05	•		20-49	18-49	-	15-49
Tanzania 2010	•	•	20-49	18-49	20-49	15-49
Tanzania 2015-16	•	•	20-49	17-49	20-49	15-49
Timor-Leste 2009-10	•	•	20-49	18-49	20-49	15-49
Togo 2013-14	•	•	20-59	16-59	20-49	15-49

Survey	Women asked to state their husband's age	Men asked to state their wife's age	Men's age range		Women's age range	
			For heaping analysis	For all other analysis	For heaping analysis	For all other analysis
Uganda 2000-01	•		25-54	16-54	-	16-49
Uganda 2006	•	•	25-54	18-54	20-49	15-49
Uganda 2011	•	•	25-54	18-54	20-49	15-49
Ukraine 2007	•		20-49	18-49	-	16-49
Zambia 2001-02	•		21-59	17-59	-	15-49
Zambia 2007	•	•	20-59	16-59	20-49	15-49
Zambia 2013-14	•	•	20-59	17-59	20-49	15-49
Zimbabwe 2005-06	•	•	25-54	18-54	20-49	15-49
Zimbabwe 2010-11	•	•	25-54	18-54	20-49	15-49
Zimbabwe 2015	•	•	25-54	16-54	20-49	15-49