USAID
FROM THE AMERICAN PEOPLE

THE EFFECT OF INTERVIEWER CHARACTERISTICS ON DATA QUALITY IN DHS SURVEYS

DHS METHODOLOGICAL REPORTS 24

SEPTEMBER 2018

This publication was produced for review by the United States Agency for International Development. It was prepared by Thomas W. Pullum, Christina Juan, Nizam Khan, and Sarah Staveteig.

DHS Methodological Reports No. 24

The Effect of Interviewer Characteristics on Data Quality in DHS Surveys

Thomas W. Pullum[1,2]
Christina Juan[1,2]
Nizam Khan[1]
Sarah Staveteig[2,3]

ICF
Rockville, Maryland, USA

September 2018

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

Corresponding author: Thomas W. Pullum, International Health and Development, ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA; phone: +1 301-572-0950; fax: +1 301-572-0999; email: tom.pullum@icf.com

Acknowledgments: The authors wish to thank Sunita Kishor and survey managers on the DHS staff for valuable insights into the interview process. We thank Yoonjoung Choi for very helpful comments on an earlier draft.

Editor: Bryant Robey
Document Production: Joan Wardell

This study was carried out with support provided by the United States Agency for International Development (USAID) through The DHS Program (#AID-OAA-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. Additional information about The DHS Program can be obtained from ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850 USA; telephone: +1 301-572-0200, fax: +1 301-572-0999, email: info@DHSprogram.com, internet: www.DHSprogram.com.

Recommended citation:

Pullum, Thomas W., Christina Juan, Nizam Khan, and Sarah Staveteig. 2018. *The Effect of Interviewer Characteristics on Data Quality in DHS Surveys*. DHS Methodological Reports No. 24. Rockville, Maryland, USA: ICF.

CONTENTS

TABLES AND LISTS

FIGURES

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

As part of a continuous effort to maintain and improve the quality of data in DHS surveys, this report examines whether variation in 25 indicators of data quality, in 15 recent DHS surveys, can be attributed to interviewers and their characteristics. The analysis is based on interviewer ID codes that appear at several points in DHS data files, and information about the interviewers obtained in a Fieldworker Survey that is now a standard component of all DHS surveys. All of the data files are publicly available.

The 25 indicators are in three broad categories: nonresponse and refusals; reported age at death of young children; and ages and dates. The third group includes five subgroups or domains: incompleteness of age, which usually takes the form of a missing month of birth; inconsistency between age in the household survey and age in the individual surveys of women or men; heaping on ages that end in 0 or 5; displacement of age across boundaries for eligibility; and a new indirect indicator of over-dispersion of children's age derived from flagging of the height-for-age and weight-for-age scores. All indicators are defined at the level of the individual, with outcome "1" for a problematic or potentially problematic response, and otherwise either "0" or "Not Applicable". Because the outcomes are binary, they can be easily analyzed with logit regression and related versions of generalized linear models. Combinations of indicators and surveys are judged to be problematic if the level or prevalence of the outcome "1" is relatively far from an acceptable level and there is highly significant variation in the outcome across interviewers. Many such combinations are identified, with systematic in-depth investigation of several examples. It is found that when there is a high degree of variation across interviewers, in terms of a data quality indicator, the bulk of that variation can often be traced to a handful of interviewers on the same team or on different teams.

To investigate the potential effect of the covariates in the Fieldworker Survey, similar indicators are pooled and all the surveys are pooled. There are exceptions, but it is generally found that interviewers who are older and better educated have lower levels of problematic outcomes. Prior experience with a DHS survey or with other surveys is often statistically significant, and often—but not always—in the direction of better quality data. There is concern when previous experience may lead to worse, rather than better, data.

The most important limitation is that interviewer assignments are almost always to just one or two geographic regions within a country, and the quality of the data they collect is confounded with potentially relevant characteristics of the regions and the composition of potentially relevant characteristics of the respondents. For example, the respondents' level of education is associated with the accuracy of their stated age, and interviewers assigned to a region with a low level of education cannot be expected to obtain the same quality of responses as interviewers who are assigned to other regions.

Further analysis is planned that will include characteristics of the respondents along with those of the interviewer, and possible statistical interactions that reflect the social distance between interviewers and respondents. The methods and findings of this study are relevant to ongoing efforts to improve the training of interviewers and the monitoring of fieldwork.

KEY WORDS: data quality, interviewer effects, interviewer training, interviewer supervision, fieldwork monitoring, nonresponse, omission of deaths, age and date reporting

1 BACKGROUND

The Demographic and Health Surveys (DHS) Program has gained a worldwide reputation for its contributions to providing invaluable data on such issues as modern contraceptive use, malaria, child health, domestic violence, women's empowerment, and biomarkers such as anemia and HIV/AIDS. Because of the value of DHS data—collected in low- and middle-income countries (LMICs) with cross-sectional, nationally representative surveys—it is important to focus continuously on measuring the quality of data and on the possibilities for improving survey methodology.

A deeper understanding of how data collection affects data quality is crucial. DHS questionnaires continue to evolve and grow, placing constraints on resources and time for the staff who support data collection and preparation (e.g., survey managers, interviewers, data processors). Over the last few years, survey questionnaires have shifted from being administered on paper to an entirely digital methodology—Computer-Assisted Personal Interview (CAPI). These two approaches to data collection data have different implications for data quality.

This report is concerned with the effect that interviewers have on data quality. How much does the quality of data vary from one interviewer to another? How do the characteristics of interviewers affect the quality of the data they collect?

1.1 Related Research on Interviewer Effects

Methodological studies of data quality have been carried out for DHS surveys and for surveys more generally. We will review some of the issues these studies have raised. A distinction can be made between two types of interviewer effects: role-restricted, referring to possible influences on survey responses of an interviewer's behavior and conduct; and role-independent, referring to background characteristics of an interviewer such as age, race, and gender (Anglewicz 2009).

When collecting survey data, several measurement issues can potentially occur. For example, the interview experience is different for respondents in face-to-face surveys and in Computer-Assisted Telephone Interviewing (CATI) surveys (Davis et al. 2009). Issues arise from social desirability bias (Davis et al. 2009; Ford and Norris 1997; Liu 2016; Wang et al. 2013; Yang and Yu 2008), and under-reporting, such as omission of early births to very young adolescents (Neal et al. 2016). These measurement errors may be influenced by the age of the interviewer (Cleary, Mechanic, and Weiss 1981; Ford and Norris 1997).

The race and ethnicity of the interviewer and the respondent also may affect measurement errors. Studies have found better data quality for non-coethnic interviewer and respondent pairs in various regions of the world (Adida et al. 2016; Davis et al. 2009; Liu 2016; Wang et al. 2013). In an examination of the effects of perceived race and ethnicity of an interviewer through telephone interviews, using the 2008 National Asian American Survey (NAAS), Liu (2016) found that when Asian American respondents perceived the interviewer to be non-Asian, they were more likely to share their experiences of discrimination.

Additionally, in a study of the interviewer-respondent dynamic in the African context, Adida et al. (2016) examined the role of ethnicity in surveys in 14 African countries. Being interviewed by someone of different ethnicity was strongly associated with differences in the responses, suggesting that the ethnicity of the interviewer and respondent, including perceived beliefs about ethnicity, are important considerations. In a study by Wang et al. (2013), Hispanic respondents in the United States were less likely to report substance use to Hispanic interviewers compared with non-Hispanic interviewers.

Studies have had mixed findings concerning how the sex of the interviewer and the respondent affects responses. In one study, female respondents were more likely to share feminist views with male interviewers (Landis et al. 1973). Another more recent study found no strong statistical evidence of an interaction between interviewer and respondent gender for gender-related variables included in the study; however, similar to the study by Landis (1973), women were somewhat more likely than men to share feminist responses with female interviewers (Huddy et al. 1997). Additionally, female interviewers felt that their respondents were more direct than male interviewers felt about their respondents (Davis et al. 2009).

In addition to role-independent interviewer characteristics such as race or sex (Bignami-Van Assche, Reniers, and Weinreb 2003), researchers have examined the potential influences of role-restricted characteristics related to an interviewer's behavior and conduct (Anglewicz 2009), such as interviewer workload, interview length, reporting of expenditures (Nix 2014), and rapport (Cleary, Mechanic, and Weiss 1981; Hox, De Leeuw, and Kreft 1991; Yang and Yu 2008). A conceptual framework has been developed to help examine different dimensions of interviewer characteristics, including interviewer attitudes, behavior, experience with measurements (such as whether they would consent to giving blood if asked in a survey), and expectations (Blom 2013). Hox et al. (1991) found mixed results regarding interviewer rapport with the respondent. One study found evidence of improved data when an interviewer is detail-oriented and exudes self-confidence. Another showed that qualities of cooperativeness were negatively associated with high-quality data. An interviewer's appearance, behavior, and perceived educational background may affect responses. One study found that respondents with less education provided more detailed responses than more highly educated respondents, trying to conform to the expectations and perceived educational background of the interviewer (Yang and Yu 2008). Another study found that older and more experienced interviewers create a relaxed and welcoming atmosphere, helping respondents feel more at ease with sharing their personal experiences (Cleary, Mechanic, and Weiss 1981).

1.2 Previous DHS Research on Data Quality

Self-reported ages and dates are prone to bias and inaccuracy. Specific to the DHS experience, age-related measurement error may include age heaping, as a result of age rounding, and age vanity—the tendency for respondents to report themselves as somewhat older in youth, somewhat younger in middle age, and considerably older in old age (Howell 1986).

The previous DHS study to which this report is most closely related is Analytical Study 19 (Johnson et al. 2009). That study investigated fieldwork-related characteristics, including the interviewer's sex, the number of visits to the household, whether a translator was present during the interview, and whether the respondent's mother tongue was the same as the language used in the interview. The data quality indicators were age heaping, missing or incomplete data, age displacement, inconsistencies in age and date reporting, and nonresponse. The study found that in most cases the interviewer's knowledge of the local language, the use of a questionnaire translated into the local language, and the timing of the interview during fieldwork (near the beginning or end of fieldwork) had important implications for data quality. In many instances, data quality was associated—either positively or negatively—with the number of visits to the household.

For example, in some countries age heaping increased with the number of visits to the household, which contrasts with findings by Howell (1986) on improved data quality if more time is spent with a survey population. Some differences in data quality were found to be associated with the number of days spent in the survey cluster doing fieldwork; however, the differences were inconsistent in direction.

Johnson et al. (2009) noted for most countries an increased likelihood of age displacement of male household members outside the ages of eligibility, but not a corresponding displacement of female household members, which denotes a reduced workload for male interviewers in countries where interviewers and respondents are matched by same sex for DHS questionnaires. For several countries they found downward age displacement among women—women age 50-54 were misreported as age 45-49—which was more likely if the interview was conducted toward the end of fieldwork. Additionally, male interviewers tended to rush through household interviews, especially as the number of visits to the household increased, making age displacement and nonresponse more likely.

The study by Johnson and colleagues also found other factors such as residence and language to be associated with data quality. Urban-rural residence was correlated with downward age displacement, and was higher in rural areas. Language differences between the interviewer and respondent were associated with greater age displacement. There was no strong or clear relationship between the time of day of the interview and the frequency of downward age displacement for women (Johnson et al. 2009).

Another DHS study (Pullum and Staveteig 2017) examined the ages of household members; self-reported ages and birthdates of women and men; women's self-reports of age and date of first union; birthdates of children, provided by the mother; and women's and men's estimates of their respective spouses' ages in the surveys of women and men. Here, too, there was evidence of age displacement and heaping.

Related to the reporting of ages and dates, data on birth history are associated with forward displacement of the date of first birth, or omission of early births, among older women, especially in sub-Saharan Africa (Blanc and Rutenberg 1990). Using the DHS survey of Bangladesh in 1994 and Matlab's Demographic Surveillance System to compare two types of maternity histories (birth histories and pregnancy histories), missing information about date of birth is more frequent among rural than among urban women (Johnson et al. 2009), and in cases where translation is needed (Espeut and Becker 2015). Espeut and Becker (2015) examined the reporting of live births and infant deaths and identified two types of errors that affect data quality: event omissions, which leads to an underestimation of demographic indicators; and event displacement, both backward and forward.

In meetings with DHS survey managers, we discussed the DHS interview process in several countries to help identify potential opportunities for bias and measurement error. Typically, a survey team of five or six people (e.g., team supervisor, driver, several female interviewers, and a biomarker specialist) are deployed to the clusters where they are to conduct the interviews. In some countries questionnaires are administered in several languages. It was noted that an interviewer only rarely asks questions about both age and birthdate. Typically, the interviewer uses a calculator and the respondent's stated age to calculate birth year. Age heaping takes place in countries because, for instance, some women are embarrassed by their age or do not know their actual age. Interviewers may be instructed to estimate the woman's age based on the number and ages of her children. In Ghana, a form of identification (e.g., a birth certificate) is requested to verify age and birthdate. Besides role-independent interviewer effects, other effects that DHS survey managers mentioned were language and class. When the questionnaire must be translated into several languages, interviewers sometimes modify the translation during the interview, believing that this will improve the understanding of the questions. Instead of reading questions verbatim—using formal terminology for a sexual behavior, for instance—the interviewer may use a slang term, promoting comfort and ease during a discussion of sensitive topics during the interview, but also producing some variability across interviewers.

1.3 Objectives and Structure of This Report

The DHS Program has always recognized the vital importance of interviewers for the collection of high-quality data. The selection, training, and supervision of interviewers follow detailed guidelines. Interviewers are acknowledged by name in the survey reports, are provided with appropriate benefits and compensation, and are also an important component of each survey's contribution to strengthening national capacity to do future surveys. Data collection can be an intense and demanding activity spread over an interval of 4-5 months, and sometimes longer.

DHS surveys include identification (ID) codes for interviewers in the data files, coming from the household, women's, and men's questionnaires. Codes for supervisors and data entry staff (when relevant) are also included. However, there has been relatively little analysis of these codes after the construction of the standard recode files, for example to relate interviewers to variations in data quality. The principal example of such an analysis conducted within DHS is the 2009 report by Kiersten Johnson and colleagues (Johnson et al. 2009) described above.

Since the beginning of DHS-7, some basic information about interviewers has been collected, using a standard format, and interviewer files have recently become available for analysis. The impetus for instituting surveys of interviewers was a specific need to better understand how interviewers may affect refusals to consent to HIV testing, and a more general need to assess whether and how interviewers affect DHS data quality. A better understanding of such effects could lead to better training and monitoring and to improved data quality in DHS surveys.

This report represents the first systematic analysis of that information, using interviewer data from 15 DHS surveys. The report has the following principal objectives:

- To describe variation across surveys in the levels of a set of indicators of data quality
- To describe variation across interviewers, within surveys, in the level of data quality
- To develop and apply procedures to identify problematic interviewers
- To describe how such variation may be related to the characteristics of the interviewers, such as their level of education and prior survey experience

Chapter 2 will describe the interviewer surveys, including the characteristics of the interviewers in the available surveys and interview procedures in the women's survey and household survey. Chapter 3 describes the methods for linking the main DHS survey data with the interviewer data and identifying interviewer effects. Chapters 4, 5, and 6 present findings for three groups of indicators: nonresponse and refusals in Chapter 4; potential misreporting of deaths in Chapter 5; and various problems with age and date reporting in Chapter 6. Chapter 7 provides conclusions and recommendations.

2 INTERVIEWER CHARACTERISTICS

2.1 Data

Table 2.1 lists the countries and surveys that included systematic collection of interviewer characteristics and are analyzed in this study. The list includes DHS data from nine countries, MIS (Malaria Indicator Survey) data from five countries, and AIS (AIDS Indicator Survey) data from one country.

Broadly speaking, these surveys included three types of interviews: an interview with the household respondent, who is usually but not necessarily the household head or the spouse of the household head, which may be conducted by either a male or a female interviewer; interviews with eligible women in the household, which are typically conducted by a female interviewer; and interviews with eligible men in the household, which are typically conducted by a male interviewer. The MIS surveys do not include a men's survey. The interviewer for individual women and men is typically the same person as the household interviewer, if the same sex.

Interviewers typically provide information about their personal characteristics at the start of fieldwork with a self-administered paper questionnaire. The information is not used for hiring or for any other purposes during fieldwork. At the end of the training, identification numbers are assigned. Each interviewer has a unique numeric ID code, which is entered in the household and individual data files and can be linked with the interviewer data.

The data used in this analysis come from four generic data files: the household data (PR file[1]); the women's data (IR file); the men's data (MR file), when applicable; and the interviewer characteristics data (FW file). All demographic characteristic variables are generated from FW data, and workload variables are generated from the IR and PR data.

[1] DHS household data appear in two standard recode files: the HR file, a "wide" file with one record for each sampled household, and the PR file, a "long" file with one record for each individual in each sampled household. For our purposes, it is easier to work with the PR file, but the same information is also in the HR file.

Table 2.1 Survey types and countries included in this study

Country	Survey Year	Data collection mode
DHS		
Armenia	2015-16	PAPI
Burundi	2016-17	CAPI
Malawi	2015-16	CAPI
Myanmar	2015-16	CAFE
Nepal	2016	CAPI
Timor-Leste	2016	CAPI
Tanzania	2015-16	CAFE
Uganda	2016	CAPI
Zimbabwe	2016	CAPI
MIS		
Ghana	2016	CAPI
Kenya	2015	CAPI
Liberia	2016	PAPI
Nigeria	2015	CAPI
Sierra Leone	2016	CAPI
AIS		
Mozambique	2016	CAPI

2.2 Interviewer Characteristics and Survey Work Experience

This section describes the distribution of interviewer characteristics for the surveys included in this study. The data include some key demographic characteristics (sex, age, marital status, whether the interviewer has living children, has had any child deaths, level of education attained, residence, ethnicity, religion, and language), and previous survey work experience. We also describe interviewer workloads, including number of interviews and some characteristics of the interviews, such as number of visits, number of interviews completed per day, and average time required to complete the interviews.

Table 2.2 summarizes the demographic characteristics and survey work experience of the interviewers. In all survey countries except Ghana and Zimbabwe, there are more female interviewers than males. In Ghana and Zimbabwe the percentages of male and female interviewers are almost the same. The higher percentage of female interviewers in most countries is explained by the fact that the questionnaires for women are longest and interviews with women are conducted almost exclusively by female interviewers.

Table 2.2 Percentage distribution of interviewers by interviewer characteristics and country

	DHS									MIS					AIS
	Armenia	Burundi	Malawi	Myanmar	Nepal	Timor-Leste	Tanzania	Uganda	Zimbabwe	Ghana	Kenya	Liberia	Nigeria	Sierra Leone	Mozambique
Gender															
Male	17.3	41.1	38.2	37.4	37.4	33.3	32.5	35.1	49.2	48.9	35.0	27.2	24.8	39.7	36.6
Female	82.7	58.9	61.8	62.6	62.6	66.7	67.5	64.9	50.0	51.1	65.0	72.8	63.7	60.3	62.7
Missing	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	11.5	0.0	0.8
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Age															
18-24	3.8	3.2	49.3	46.3	22.0	19.2	3.2	9.9	15.0	6.6	29.3	1.7	4.6	2.7	12.3
25-29	3.8	25.3	23.3	26.5	45.1	39.2	20.6	47.0	45.0	44.3	18.8	23.3	18.5	24.8	31.3
30-34	3.8	26.6	7.8	4.1	18.7	22.5	11.1	21.2	20.0	18.0	8.3	46.7	19.0	28.3	34.1
35-39	8.7	22.8	6.4	2.7	9.9	4.2	13.5	8.0	5.8	14.8	9.8	16.7	18.5	16.8	10.6
40-44	12.5	12.7	2.0	6.1	2.2	4.2	23.8	5.3	3.3	6.6	12.0	8.3	11.3	8.9	7.3
45-49	21.2	5.7	2.0	5.4	2.2	1.7	12.7	3.3	3.3	6.6	12.0	3.3	16.9	5.3	1.1
50+	46.2	3.8	2.7	6.8	0.0	0.0	15.1	1.3	0.8	3.3	9.8	0.0	10.3	1.8	2.2
Missing	0.0	0.0	6.4	2.0	0.0	9.2	0.0	4.0	6.7	0.0	0.0	0.0	1.0	11.5	1.1
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Mean	47.8	34.6	26.7	29.0	28.4	28.3	38.5	30.2	29.6	32.5	33.5	32.8	37.5	33.4	31.2
Marital Status															
Currently married	78.8	59.5	28.4	21.8	41.8	40.8	70.6	39.1	33.3	37.7	52.6	53.3	66.7	43.4	52.5
Previously married	9.6	1.3	1.4	1.4	1.1	2.5	29.4	5.3	5.0	0.0	6.0	3.3	5.6	8.9	12.9
Never married	11.5	39.2	63.9	74.8	57.1	49.2	0.0	51.7	57.5	60.7	41.4	43.3	26.7	36.3	31.8
Missing	0.0	0.0	6.4	2.0	0.0	7.5	0.0	4.0	4.2	1.6	0.0	0.0	1.0	11.5	2.8
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Highest level of education attended															
Attended or completed secondary education	42.3	15.2	27.7	100.0	98.9	19.2	100.0	0.7	3.3	3.3	2.3	3.3	1.0	5.3	51.4
Attended post-secondary education	57.7	84.8	65.9	0.0	1.1	73.3	0.0	95.4	92.5	96.7	97.7	96.7	98.5	83.2	48.6
Missing	0.0	0.0	6.4	0.0	0.0	7.5	0.0	4.0	4.2	0.0	0.0	0.0	0.5	11.5	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Number of kids															
0	12.5	43.0	66.9	84.4	73.6	56.7	30.2	50.3	54.2	65.6	42.9	16.7	32.8	21.2	24.0
1	15.4	7.6	11.5	3.4	14.3	16.7	18.3	23.8	20.0	8.2	16.5	38.3	14.4	30.1	29.1
2	46.2	17.7	7.1	6.8	11.0	5.8	23.0	13.3	12.5	6.6	14.3	25.0	9.7	23.9	25.1
3	19.2	13.9	3.4	2.0	1.1	5.8	8.7	6.6	4.2	9.8	11.3	15.0	18.0	9.7	12.3
4+	6.7	17.7	4.7	1.4	0.0	12.5	19.8	2.0	1.7	8.2	15.0	5.0	23.1	3.5	7.3
Missing	0.0	0.0	6.4	2.0	0.0	2.5	0.0	4.0	7.5	1.6	0.0	0.0	2.1	11.5	2.2
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Mean	1.9	1.6	0.6	0.4	0.4	1.4	1.8	0.8	0.7	2.5	1.6	1.6	2.1	1.4	1.5
Experienced any child death															
Yes	10.6	4.4	3.0	1.4	0.0	4.2	6.4	3.3	4.2	0.0	6.0	5.0	15.9	9.7	12.3
No	89.4	95.6	90.5	96.6	100.0	86.7	93.7	92.7	90.0	95.1	94.0	95.0	82.1	78.8	82.1
Missing	0.0	0.0	6.4	2.0	0.0	9.2	0.0	4.0	5.8	4.9	0.0	0.0	2.1	11.5	5.6
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

(Continued…)

Table 2.2—*Continued*

	DHS									MIS					AIS
	Armenia	Burundi	Malawi	Myanmar	Nepal	Timor-Leste	Tanzania	Uganda	Zimbabwe	Ghana	Kenya	Liberia	Nigeria	Sierra Leone	Mozambique
Residence															
City	35.6	77.9	50.3	-	-	62.5	27.0	32.4	-	59.0	29.3	96.7	61.5	76.1	84.9
Town	48.1	12.0	24.7	-	-	19.2	51.6	53.0	-	36.1	45.9	1.7	33.3	11.5	6.7
Urban	-	-	-	51.7	83.5	-	-	-	85.9	-	-	-	-	-	-
Rural	16.3	10.1	18.6	46.3	16.5	10.8	21.4	10.6	8.3	1.6	24.8	1.6	4.6	0.9	8.4
Missing	0.0	0.0	6.4	2.0	0.0	7.5	0.0	4.0	5.8	3.3	0.0	0.0	0.5	11.5	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Work experience															
Worked in DHS/MIS	32.7	25.3	12.2	18.4	13.2	10.8	34.9	17.2	24.2	44.3	21.8	15.0	38.0	41.6	24.6
Worked in other survey	50.0	57.6	30.1	21.8	62.6	41.7	42.1	53.6	37.5	41.0	39.9	70.0	30.8	31.9	41.3
No prior experience	17.3	17.1	51.4	57.1	24.2	40.0	23.0	25.2	31.7	13.1	38.4	15.0	29.2	15.0	30.2
Missing	0.0	0.0	6.4	2.7	0.0	7.5	0.0	4.0	6.7	1.6	0.0	0.0	2.1	11.5	3.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Total number of Interviewers	**104**	**158**	**296**	**147**	**91**	120.0	126.0	**151**	**120**	**61**	**133**	**60**	**195**	**113**	**179**

There are wide variations in interviewer age across countries. The percentage of young interviewers (age 18-24) is relatively high in Malawi (49%) and Myanmar (46%), followed by Kenya (29%), and Nepal (22%). In contrast, the percentage of older interviewers (age 40 and older) is overwhelmingly high in Armenia (80%), followed by Nigeria (39%) and Kenya (34%). In Armenia most of the interviewers are age 50 and older.

In most surveys a majority of interviewers are ever-married. The percentage never-married is highest in Myanmar (75%), followed by Malawi (64%), Zimbabwe (58%), and Nepal (57%).

The self-administered interviewer questionnaire asked about the highest level of education attended. All interviewers attended at least secondary level. In Burundi, Ghana, Kenya, Liberia, Nigeria, Sierra Leone, Uganda, and Zimbabwe, more than 80% had completed secondary education and had some post-secondary education. In Myanmar all interviewers and in Nepal 99% of interviewers had attended or completed secondary education.

The interviewer questionnaire included questions on having living children and on experience with any child deaths, irrespective of marital status. Interviewers who have children and those who experienced child deaths may be able to administer survey questions about respondents' children's births and deaths with more sensitivity.

In five countries a majority of the interviewers said they had no children. More than three-fourths of the interviewers in Myanmar (84%) and nearly three-fourths in Nepal (74%) had no living children, while about two-thirds of the interviewers in Malawi (67%) and in Ghana (66%) reported having no living children. Slightly more than half of the interviewers in Zimbabwe (54%) and half of the interviewers in Uganda (50%) had no children. In contrast, 40% of interviewers in Nigeria, 32% in Burundi, and 26% in Kenya and Armenia had three or more children.

Few interviewers had experienced any child deaths. The percentage was highest in Nigeria (16%), followed by Mozambique (12%), Armenia (11%), and Sierra Leone (10%). The relatively high percentage for Armenia is somewhat surprising because that country has a low level of child mortality, but may be due to the older age distribution of interviewers.

Most of the interviewers lived in cities or towns. The highest percentage of interviewers who lived in rural areas was in Myanmar (46%), followed by Kenya (25%).

Tables in Appendix 1 show the distribution of interviewers' ethnicity, religion, and mother tongue, all of which are survey specific.

2.3 Interviewer Workloads in the Women's Survey

The women's interview, which is administered to all eligible women age 15-49 identified in the household survey, is substantially longer for a DHS survey than for an MIS or AIS survey. In a DHS survey the questionnaire includes questions on women's health, nutrition, empowerment, gender violence, child health and nutrition, and a complete birth history. MIS and AIS questionnaires for women include fewer questions about women and children, and only a truncated birth history for births in the past 5 years. Table 2.3 shows the length of fieldwork and the total number of interviewers and interviews for the women's survey by country.

Table 2.3 Length of fieldwork and total number of interviewers and interviews by country, for the survey of women

	Length of fieldwork in days	Total number of interviewers	Total number of women interviewed
DHS			
Armenia	115	52	6,116
Burundi	146	76	17,269
Malawi	118	148	24,562
Myanmar	213	79	12,885
Nepal	227	57	12,862
Timor-Leste	107	80	12,607
Tanzania	176	111	13,266
Uganda	182	91	18,506
Zimbabwe	166	60	9,955
MIS			
Ghana	65	37	5,150
Kenya	39	52	5,394
Liberia	66	29	4,290
Nigeria	40	141	8,034
Sierra Leone	36	80	8,501
AIS			
Mozambique	169	78	15,498

Table 2.4 Percentage distribution of interviewers by workload and country (Women's Survey)

	DHS							MIS					AIS
	Armenia	Burundi	Malawi	Myan-mar	Nepal	Uganda	Zimba-bwe	Ghana	Kenya	Liberia	Nigeria	Sierra Leone	Mozam-bique
Number of visits													
1	94.2	69.7	70.3	97.5	68.4	78.0	66.7	86.5	80.8	93.1	89.4	87.5	96.2
2	5.8	25.0	21.0	2.5	26.3	17.6	21.7	10.8	15.4	3.5	9.9	7.5	3.9
3	0.0	2.6	4.7	0.0	3.5	4.4	6.7	2.7	1.9	3.5	0.0	3.8	0.0
4	0.0	1.3	2.7	0.0	0.0	0.0	1.7	0.0	1.9	0.0	0.0	1.3	0.0
5	0.0	1.3	1.4	0.0	1.8	0.0	3.3	0.0	0.0	0.0	0.7	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Overall mean	1.1	1.4	1.4	1.0	1.4	1.3	1.5	1.2	1.3	1.1	1.1	1.2	1.0
Number of interviews completed by an Interviewer per day													
1	34.6	51.3	18.9	35.4	59.7	36.3	40.0	10.8	13.5	13.8	20.6	2.5	30.8
2	42.3	36.8	29.1	36.7	35.1	30.8	20.0	5.4	1.9	6.9	25.5	20.0	32.1
3	9.6	7.9	23.7	15.2	5.3	20.9	26.7	29.7	15.4	34.5	21.3	15.0	14.1
4	5.8	4.0	17.6	10.1	0.0	8.8	5.0	13.5	19.2	6.9	9.2	11.3	12.8
5	5.8	0.0	9.5	2.5	0.0	2.2	5.0	13.5	15.4	13.8	12.1	21.3	6.4
6+	1.9	0.0	1.4	0.0	0.0	1.1	3.3	27.0	34.6	24.1	11.4	30.0	3.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Overall mean	2.1	1.6	2.7	2.1	1.5	2.1	2.3	4.2	4.6	3.9	3.2	4.8	2.4
Total number of interviews completed by an Interviewer during the survey period													
<100	36.5	17.1	11.5	24.1	8.8	4.4	26.7	5.4	28.9	17.2	95.7	36.3	41.0
100-149	44.2	0.0	22.3	20.3	5.3	15.4	13.3	56.8	71.2	17.2	4.3	47.5	39.7
150-199	19.2	5.3	41.2	21.5	12.3	24.2	20.0	35.1	0.0	44.8	0.0	15.0	19.2
200-249	0.0	14.5	18.2	16.5	45.6	36.3	23.3	2.7	0.0	20.7	0.0	1.3	0.0
250+	0.0	63.2	6.8	17.7	28.1	19.8	16.7	0.0	0.0	0.0	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Overall mean	117.5	227.2	165.9	163.0	225.6	203.4	165.7	139.0	103.7	147.7	56.9	106.2	99.1
Number of interviewers	**52**	**76**	**148**	**79**	**57**	**91**	**60**	**37**	**52**	**29**	**141**	**80**	**78**
Average time to complete the interview (in minutes)													
<20	0.0	0.0	1.0	0.0	0.0	0.0	0.0	65.6	92.9	22.2	53.2	68.6	6.7
20-39	16.3	1.9	10.6	11.7	0.0	2.8	15.0	31.3	7.1	44.4	43.7	30.0	25.3
40-59	32.7	9.4	25.0	24.7	20.5	11.3	42.5	3.1	0.0	18.5	2.4	0.0	44.0
60-79	26.5	15.1	31.7	23.4	28.2	14.1	22.5	0.0	0.0	11.1	0.8	0.0	17.3
80-99	14.3	17.0	22.1	11.7	12.8	25.4	17.5	0.0	0.0	0.0	0.0	1.4	5.3
100-119	6.1	22.6	6.7	19.5	12.8	25.4	2.5	0.0	0.0	0.0	0.0	0.0	0.0
120+	4.1	34.0	2.9	9.1	25.6	21.1	0.0	0.0	0.0	3.7	0.0	0.0	1.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Overall average	**55.4**	60.1	43.9	51.3	71.2	58.2	41.6	13.5	7.3	28.0	13.3	12.3	39.7
Number of interviewers (Includes cases completed in one visit)	**49**	**53**	**104**	**77**	**39**	**71**	**40**	**32**	**42**	**27**	**126**	**70**	**75**

Table 2.4 presents the distribution of the workloads of interviewers who conducted the survey of women. The data presented here come from the women's (IR) files. Interviewers who completed fewer than five interviews are excluded.

In all but two countries, about 70% or more of the women's interviews were completed during a single visit. The exception was Timor-Leste (51%).

The average number of women's interviews completed per day varies from 2.1 (Armenia) to 4.6 (Kenya and Sierra Leone). More than half of the interviewers in Nepal, and half of the interviewers in Burundi, completed only one interview per day. By contrast, about one-third of interviewers in Kenya and Sierra Leone, and about one-fourth in Ghana and Liberia, completed an average of six or more interviews per day. Variations in the length of the questionnaires and variations in terrain types and transportation in the survey areas could explain the variation in these numbers.

The total number of workdays of the interviewers depends on the sample size, proximity of sample units, type of transportation, and type of terrain. Interviewers who worked for longer periods may feel fatigued at the later stage of the survey, which may affect data quality.

On average, interviewers in Nepal completed the most interviews (226), followed by Burundi (227), Uganda (203), Malawi (166), and Myanmar (163). On average, interviewers in Nigeria completed the fewest (57), followed by Sierra Leone (106).

The bottom panel of Table 2.4 examines the time (in minutes) needed to complete the women's survey. Overall, as would be expected because of questionnaire length, DHS surveys had a longer interview time than MIS and AIS surveys. Start time and end time for interviews that needed more than one visit to complete are excluded from the calculation of interview length.

2.4 Interviewer Workloads in the Household Survey

Table 2.5 shows the length of fieldwork and the total number of interviewers and interviews for the household survey in each country.

Table 2.5 Length of fieldwork and total number of interviewers and interviews by country, for the household survey

	Length of fieldwork in days	Total number of interviewers	Total number of households interviewed
DHS			
Armenia	119	65	7,893
Burundi	146	113	15,977
Malawi	118	212	26,361
Myanmar	219	125	12,500
Nepal	226	75	11,040
Timor-Leste	117	112	11,502
Tanzania	407	112	12,563
Uganda	69	122	5,345
Zimbabwe	167	114	10,534
MIS			
Ghana	65	37	5,841
Kenya	39	52	6,481
Liberia	80	28	4,218
Nigeria	59	141	7,745
Sierra Leone	36	78	6,719
AIS			
Mozambique	141	104	7,169

Table 2.6 presents the distribution of workloads of the interviewers who conducted the household survey. The household survey includes the household roster and questions on other household characteristics such as housing, source of water, type of sanitation, etc. The data provided here come from the household (PR) file. Interviewers who completed fewer than five interviews are excluded from analysis.

Table 2.6 Percentage distribution of interviewers by work load and country (Household Survey)

	DHS							MIS					AIS
	Armenia	Burundi	Malawi	Myan-mar	Nepal	Uganda	Zimba-bwe	Ghana	Kenya	Liberia	Nigeria	Sierra Leone	Mozam-bique
Number of visits													
1	100.0	86.7	90.1	100.0	92.0	95.1	79.8	97.3	78.9	100.0	82.3	93.6	95.2
2	0.0	13.3	7.6	0.0	6.7	3.3	13.2	2.7	13.5	0.0	12.8	6.4	4.8
3	0.0	0.0	1.9	0.0	1.3	1.6	2.6	0.0	5.8	0.0	4.3	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	1.9	0.0	0.0	0.0	0.0
5	0.0	0.0	0.5	0.0	0.0	0.0	3.5	0.0	0.0	0.0	0.7	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Overall mean	1.0	1.1	1.1	1.0	1.1	1.1	1.4	1.0	1.3	1.0	1.2	1.1	1.0
Number of interviews completed by an Interviewer per day													
1	21.5	47.8	12.3	52.8	57.3	11.5	44.7	0.0	5.8	7.1	18.4	7.7	55.8
2	29.2	38.9	31.6	31.2	26.7	4.9	32.5	2.7	5.8	14.3	28.4	24.4	19.2
3	21.5	8.9	28.8	10.4	14.7	13.1	12.3	13.5	9.6	35.7	20.6	24.4	12.5
4	10.8	4.4	20.3	4.0	0.0	21.3	9.7	27.0	9.6	28.6	15.6	19.2	9.6
5	10.8	0.0	5.2	1.6	1.3	16.4	0.9	16.2	21.2	10.7	9.9	19.2	1.9
6+	6.2	0.0	1.9	0.0	0.0	32.8	0.0	40.5	48.1	3.6	7.1	5.1	1.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Overall mean	2.8	1.7	2.8	1.7	1.6	4.5	1.9	5.2	5.6	3.4	3.0	3.4	1.9
Total number of interviews completed by an Interviewer during the survey period													
<50	1.5	8.0	9.0	28.8	6.7	18.0	25.4	2.7	3.9	10.7	31.2	21.8	28.9
50-99	26.2	13.3	12.3	31.2	30.7	39.3	31.6	0.0	15.4	3.6	68.8	23.1	51.9
100-149	47.7	30.1	50.5	16.8	36.0	42.6	30.7	2.7	67.3	3.6	0.0	55.1	19.2
150-199	24.6	38.9	25.5	12.8	4.0	0.0	8.8	94.6	13.5	82.1	0.0	0.0	0.0
200+	0.0	9.7	2.8	10.4	22.7	0.0	3.5	0.0	0.0	0.0	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Overall mean	121.4	141.5	124.7	99.8	147.2	87.6	92.4	157.5	124.6	150.4	55.3	86.0	68.5
Number of interviewers	**65**	**113**	**212**	**125**	**75**	**122**	**114**	**37**	**52**	**28**	**141**	**78**	**104**
Average time taken to complete the survey (in minutes)													
<10	1.5	0.0	1.6	N/A	0.0	3.5	1.1	8.3	7.3	0.0	4.3	1.4	N/A
10-19	40.0	1.0	18.3		1.5	34.5	29.7	55.6	41.5	32.1	23.3	31.5	
20-29	56.9	71.4	77.5		88.4	56.9	67.0	36.1	48.8	64.3	67.2	63.0	
30+	1.5	27.6	2.6		10.1	0.0	1.1	0.0	2.4	3.6	5.2	4.1	
Missing	0.0	0.0	0.0		0.0	5.2	1.1	0.0	0.0	0.0	0.0	0.0	
Total	100.0	100.0	100.0		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
Overall average	**20.0**	27.8	19.9		27.0	19.4	18.2	16.4	14.1	18.4	20.9	17.7	
Number of interviewers (Include cases completed in one visit)	**65**	**98**	**191**	**125**	**69**	**116**	**91**	**36**	**41**	**28**	**116**	**73**	**99**

Note: Survey start and end time for interviews that needed more than one visit to complete have some problems and thus are excluded for time calculation.
N/A = Not available

In all countries except Zimbabwe and Kenya, more than 80% of the interviews for the household survey were completed in the first visit. The average number of household interviews completed by an interviewer varies from 1.5 (Timor-Leste) to 5.6 (Kenya). While more than half of the interviewers in Timor-Leste, Nepal, Mozambique, and Myanmar completed only one interview per day, substantial percentages of interviewers in Kenya, Ghana, and Uganda completed an average of six or more interviews per day.

On average, interviewers in Ghana completed the most household interviews (158) during the survey period, followed by Liberia (150), and Nepal (147). On average, interviewers in Nigeria completed the fewest (55), followed by Mozambique (69) and Sierra Leone (86). The last panel of Table 2.6 examines the time (in minutes) needed to complete the household survey. Most of the household interviews were completed in an average of about 20 minutes.

3 METHODS TO IDENTIFY INTERVIEWER EFFECTS

This report is methodological in two senses. First, as an assessment of data quality, it will analyze data from the available national household surveys that have accompanying information about the interviewers, in order to improve our understanding of variations in the quality of the household data and whether those variations may be related to the interviewers. This analysis may lead to improvements in the training and supervision of interviewers. Second, the report seeks to advance the methods to assess data quality, and particularly to identify interviewer effects on data quality. With further refinement, these methods can be applied to future surveys on a routine basis after the data have been collected. These methods also have the potential to be applied during data collection, for early detection of problematic responses and identification of interviewers who have especially high levels of problematic responses, to supplement the field check tables that always accompany data collection.

Using 25 indicators, grouped into three main topics, and 15 surveys, the methods are intended to answer the following questions:

- Did the outcome occur often enough to justify further investigation?
- Was there significant[2] variation, across interviewers, in the prevalence of the outcome?
- Were there specific interviewers for whom the prevalence of the outcome was significantly greater than the overall prevalence?
- How concentrated was the outcome, in terms of its association with just a few specific interviewers?
- Were the problematic outcomes associated with measured characteristics of the interviewers?

All of these questions except for the last one could apply to any survey that has consistent[3] identification codes for interviewers, even if there is no accompanying information about the characteristics of the interviewers.

Investigations of data quality typically involve the calculation of indicators that compare an observed value with an expected value. For example, digit preference can be measured by comparing the distribution across final digits 0 through 9 with a uniform distribution (using the Index of Dissimilarity) or with a distribution that has been adjusted for the shape of the age distribution (using Myers' Blended Index). The measure can be interpreted as the minimum percentage of reported cases that would have to be shifted from over-reported digits to under-reported digits in order to obtain the uniform—or blended—age distribution.

[2] "Significant" refers to a high level of statistical significance, such as a p-value less than .01 or even .001.

[3] In some older surveys, it is believed that the numeric ID codes for interviewers were sometimes shared, and did not consistently identify specific individuals.

A limitation of such measures is that they are only interpretable at an aggregate level; they do not have an individual-level interpretation. The demographic literature includes several such indicators. Often they lack any corresponding tests of significance.

In this report, all the indicators are binary, can be calculated for households or individual respondents, can be summarized for all the interviews done by every interviewer, and are amenable to statistical testing. They take the value "1" for an individual-level outcome that is undesirable, or potentially undesirable, or identifies a category with potential omission, and the value "0" otherwise, except that they will be NA (not applicable) if, say, the outcome is not assessed at all.

For convenience, the 25 binary indicators are listed below. Detailed definitions are provided in Chapters 4, 5, and 6.

Nonresponse for women
Nonresponse for men
Women's refusal of HIV testing
Men's refusal of HIV testing
Women's refusal of height, weight, and hemoglobin measurement
Heaping on death at 12 months of age
Death in the first month
Death in the first year
Incompleteness of women's age
Incompleteness of women's age at first union
Incompleteness of men's age
Incompleteness of children's age
Age change for women
Age change for men
Age change for children
Age heaping of household members
Age heaping of women
Age heaping of men
Age heaping of children in the birth histories
Women, low-end age transfers
Women, high-end age transfers
Men, low-end age transfers
Men, high-end age transfers
Birth year transfers
Over-dispersion of children's ages

The first indicator on this list will be used to illustrate how all the indicators are constructed. Suppose that a woman in the household survey is found to be a de facto resident of the household and age 15-49. Unless this is a survey of ever-married women, or a survey in which de jure residency is a criterion for eligibility for the individual interview, this woman will then have hv117=1 in the household file, meaning that she is eligible for the individual interview. If a woman with hv117=1 does not appear in the file of individual women, she will be given the value "1" for the indicator of nonresponse for the women's interview.

Otherwise, if she is interviewed, she will be given the value "0". Persons in the household survey for whom this variable is irrelevant, such as women who do not satisfy the criteria for eligibility, will be "NA".

For another example, suppose that someone in the household survey is reported to have an age ending in final digit 0 or 5. This person will be coded "1" on the indicator for age heaping. Otherwise, the person will be coded "0". In this example, the code "1" does not necessarily indicate a problem. It will be correct about 20% of the time, because 0 and 5 are two of the ten final digits possible. However, if the percentage of cases with code "1" is substantially greater than 20%, we can infer that there has been heaping or digit preference. If it is substantially less than 20%, then we can infer that the interviewer is intentionally avoiding these digits, usually as a result of an overemphasis during training and supervision on avoiding heaping.

The frequency of binary outcomes can readily be presented as percentages, and can easily be tested for significance. In this report, levels and tests will all be reported without using any information about the survey design, such as sampling weights, clustering, and stratification.[4] The purpose is not to produce unbiased population estimates. Every case will be counted equally. Results will sometimes only be described if the level of the problem exceeds some threshold—for example, if the level of nonresponse exceeds 2%.

For a specific survey and binary outcome, we need to assess the role of the interviewers. This will be done at two levels. First, we consider the question of whether there is significant variation across interviewers. If there is not, then there is no point in proceeding further, except to note that there may be across-the-board issues that affect all interviewers to essentially the same degree. To make this assessment, we perform a logit regression of the binary outcome on a categorical variable for interviewer ID code[5] and calculate the total deviance D. D has a chi-square distribution with K-1 degrees of freedom, where K is the number of interviewers. If there is significant variation at the .01 level of significance, then we continue. A summary table will present the percentage of cases in the "1" category of all indicators and surveys and the chi-square statistic for the logit regression just described.

If there is significant variation across interviewers, we then identify any problematic interviewers. This will be done by examining terms that represent the deviation of each interviewer's prevalence of the outcome from the overall prevalence, called "deviance residuals". The statistical literature on generalized linear models applies this label to two terms, one of which is the square of the other. We shall distinguish between them with the labels "dr" and "dr2". dr is the deviance residual produced by Stata, with a post-estimation command following a generalized linear model (glm), and dr2 (the square of dr) is the deviance residual defined by McCullagh and Nelder (1989, pp. 39 and 118) in a major reference work on such models. dr measures the interviewer-specific deviation from the overall level of the outcome. It will be negative if the prevalence for the interviewer is less than the overall prevalence and positive if the prevalence for the interviewer is greater than the overall prevalence. It has a unit normal ("z") sampling distribution under the

---

[4] Estimates in survey reports usually do adjust for the survey design. For that reason, there may be small differences between sample sizes and statistics in this report and corresponding numbers in the reports.

[5] As Appendix 2 describes, the method is equivalent to a fixed-effects model applied to the respondent-level files, but instead uses interviewer-level files that have been aggregated or collapsed.

null hypothesis that the probability of the outcome is the same for all interviewers. dr can be interpreted as the deviation from the overall level, adjusted for the number of interviews done by the interviewer.

The square of dr, dr2, is also calculated for each interviewer. Under the same null hypothesis, it has a chi-square distribution with one degree of freedom. The sum of dr2, across all interviewers, is the overall deviance D for the model. The overall deviance has a chi-square distribution with K-1 degrees of freedom, where K is the number of interviewers, and is used to test the overall null hypothesis that all interviewers have the same probability of the outcome. We will flag interviewers whose deviance residual is significant at the .01 or .001 level, primarily the latter. We require a high level of significance because of the large number of test statistics being produced. Details on the calculation of dr and dr2 are given in Appendix 2.

The identification of problematic interviewers will be modified for indicators that should never take the value "1". Suppose, for example, that the overall prevalence of such an outcome is 10%. It could happen that a specific interviewer is significantly below that level—having perhaps not a single problematic interview. We will not flag such interviewers. We will focus on interviewers who are significantly above the overall level, and not on those who are significantly above 0 or significantly below the overall mean.

For indicators that can legitimately take the value "1", for example, the prevalence of final digits 0 and 5, mentioned earlier, the direction of the deviation from the overall percentage is not taken into account. Some interviewers will over-report the outcome, perhaps just reflecting the respondents' lack of knowledge of their ages. However, as mentioned, interviewers sometimes overcorrect by under-reporting the specified outcome because of an emphasis during training or supervision on avoiding rounding or heaping. This type of deviation should also be flagged.

For indicators in which only a very low value is problematic, for example the identification of neonatal deaths, suggesting potential omission of such deaths, only deviations that are below the mean will be flagged.

The total deviance D is a measure of the overall variation, across interviewers, in the prevalence of problematic responses. Certainly there is a concern with interviewers who have any problematic responses, but the greatest concern is with those who provide the largest share of the total deviance. For some indicators we will describe this concentration with two additional measures, to supplement the .01 and .001 tests of significance. One of them is the Gini coefficient of concentration; the other is the number of interviewers who are responsible for half of the problematic deviance.

Whatever the value of the total deviance, it is always possible to identify interviewers who are above the mean, and usually a few interviewers account disproportionately for a large share of the deviance. However, to repeat, we will focus on combinations of surveys and indicators for which the deviance is quite high.

The interpretation of these measures will be described in more detail when they are first used, in Chapter 4. Additional details about the methods will also be described within the presentation of findings.

The covariates in the fieldworker questionnaire that will be analyzed are as follows:

Place of residence: 1 City; 2 Town; 3 Rural
Age: 1 <25; 2 25-34; 3 35-44; 4 45+
Sex: 1 Male; 2 Female
Marital status: 1 Currently married; 2 Previously married; 3 Never married
Has children: 0 No living children; 1 Any living children
Had a child death: 0 No child deaths; 1 Any child deaths
Education: 0 Lower education; 1 Higher education
DHS experience: 0 No; 1 Yes
Other survey experience: 0 No; 1 Yes
Permanent employee of the implementing agency: 0 No; 1 Yes

These covariates are taken directly from the coding in the interviewer surveys, with the exception of education, which is recoded from fw109 and fw110 and divides the overall empirical distribution for all surveys into two groups of approximately equal size. The description of education in each survey provided in Chapter 2 was more detailed. The simple dichotomy is only used for the pooled statistical analysis.

The Mozambique AIS survey will be omitted from the analysis of the covariates, because we do not have confidence in the linkage of the interviewer ID codes in the main survey with the ID codes in the interviewer survey. It appears that two numbering systems were used in this survey. We believe that the effect on the identification of problematic interviewers is small, but would be more serious for the analysis of covariates.

In Chapters 4-6 of this report, we do not discuss characteristics of the respondents, because the interest is in variations across interviewers and potentially according to characteristics of the interviewers. Future analysis may extend to include characteristics of the respondents and how they may interact with characteristics of the interviewers. When examining characteristics of the interviewers, the analysis will include only one covariate at a time. The number of interviewers for each combination of country and indicator is relatively small. The covariates are inherently categorical or have been categorized into two, three, or four categories, one or more of which may contain only a few cases. Multivariate analysis is limited by the number of degrees of freedom available.

4 NONRESPONSE AND REFUSALS

The household interview provides a good deal of information about the household as a unit, such as indicators of assets that make up the DHS Wealth index, source of water, sanitation, etc. It also provides some information about individuals in the household, particularly sex, age, and residency status. Other information, such as parental survival for children under age 18, school attendance, and biometric measurements, is also collected during the household interview. Apart from obtaining this kind of information, the major function of the household survey is to identify women eligible for the survey of women, and men eligible for the survey of men (if a men's survey is included). Eligibility is based on sex, age, and residency (usually de facto residence, i.e., "slept here last night," but sometimes de jure residence, i.e., "usual" residency).

If an adult identified as eligible is not subsequently interviewed in the women's or men's survey, the case counts as nonresponse. This type of nonresponse can occur because of refusals or because the respondent cannot be found at home, even with multiple visits (usually the maximum is three visits). The first chapter of DHS country reports typically includes a statement of the number and percentage of cases that are lost in this way. The DHS generally has very high response rates and, during interviewer training, places much emphasis on obtaining high response rates.

For surveys that include HIV testing, adults who are eligible for testing may be unavailable or may refuse to be tested. Refusal is the only reason for not being tested that will be described here. If the survey includes measurements of height and weight, and blood tests for hemoglobin concentration, for adult women, we also describe the level of refusals for those measurements.

Refusal to participate in any part of the survey is certainly an acceptable and respected option for all eligible respondents; participation is completely voluntary. Nevertheless, if an interviewer has an unusually high level of nonresponse or refusals compared with other interviewers, it is possible that the interviewer is not providing sufficient background for the respondent to make an informed decision. Interviewers are always motivated to lighten their workload, and dropping a respondent for further interviewing or testing is an efficient way to accomplish this.[6]

4.1 Indicators

This chapter includes five indicators. The first one, "women_NR", is the only one of the five that can be calculated for all surveys, including the five MIS surveys (Ghana, Kenya, Liberia, Nigeria, and Sierra Leone). The two HIV testing indicators ("women_hiv_refusal" and "men_hiv_refusal") can only be calculated for the four surveys that included HIV testing (Burundi, Malawi, Mozambique, and Zimbabwe). The final indicator (wt_ht_hb_refusal) cannot be calculated for MIS surveys or for the Mozambique AIS survey because they did not include measurements of height, weight, or hemoglobin concentration.

[6] Interviewers are normally paid per day, rather than per interview. A shorter interview is easier for the interviewer, but there is usually not a motivation to reduce the total number of days of interviewing.

Nonresponse for women: women_NR. The standard household-level DHS variable hv117 is coded 1 if the household member is eligible to be interviewed in the survey of women, based on age (hv105) in the range 15 to 49, sex=female (hv104=2), and de facto residency ("slept here last night", hv103=1). When the file of women is merged with the file of household members, there are always a few women with hv117=1 who do not appear in the file of women. There are several possible reasons. It may happen that the woman is not found at home, even after multiple visits or postponements. She may refuse. She may be incapacitated in some way. There may be "other" reasons. Some reasons, if specified correctly, are outside the control of the interviewer. We will not distinguish among these reasons. Women_NR is coded 0 if hv117=1, and is recoded to 1 if the woman is NOT in the women's data. The nonresponse rate given in DHS reports is equivalent to the proportion of cases coded 1, among those who are coded 0 or 1, multiplied by 100 to give a percentage. The nonresponse is linked to the ID code for the household interviewer.

Nonresponse for men: NR_men. A similar indicator for men is coded 0 if the man satisfies the age and residency criteria (sometimes there is subsampling of households for the men's survey), that is, hv118=1. Such cases are recoded to 1 if the man is not then found in the men's file. The nonresponse is linked to the ID code for the household interviewer.

Women's refusal of HIV testing: **women_hiv_refusal**. HIV testing was included in the Mozambique AIS survey and in three of the DHS surveys as identified above. A variable in the household file, ha63, indicates whether the woman or man was tested, or was absent at the time of testing, or refused to be tested. Using ha63, the indicator is defined to be 1 if a woman refused to be tested, and 0 otherwise, except that it is NA if the woman was not included in the survey of women. The ID code for the biomarker specialist is used.

Men's refusal of HIV testing: men_hiv_refusal. Using hb63, the indicator is defined to be 1 if a man refused to be tested, and 0 otherwise, except that it is NA if the man was not included in the survey of men. The ID code for the biomarker specialist is used.

Women's refusal of height, weight, and hemoglobin measurement: Women ht wt hb refusal. Respondents have the option of refusing the measurements of height, weight, and hemoglobin concentration. Interviewers with a high proportion of refusals may be underperforming. Note that these measurements are made by the biomarker specialist, during the household interview. This indicator is available in all the DHS surveys in this report but not the AIS and MIS surveys, because they did not include such measurements. We do not include a separate indicator for men's refusals. The indicator is 1 if a woman who otherwise was successfully interviewed has a refusal code for any combination of the measurements of height, weight, and hemoglobin concentration (that is, ha13=4 or ha55=4). It is 0 otherwise, except that it is NA if, for example, the survey did not include these measurements for women age 15-49. The ID code for the biomarker specialist is used.

4.2 Identifying Problematic Surveys and Indicators

This report uses terms such as "problems" and "problematic" to refer generically to outcomes that are clearly undesirable in themselves, such as nonresponse, or are potential errors, such as over-reporting ages that end in the digits "0" or "5". Such outcomes will not be an issue unless they occur much more often or much less often than expected.

If the level of a problematic outcome is uniform across interviewers, or nearly uniform—whether uniformly low or uniformly high—then there are no “interviewer effects.” If the overall level is considered to be too low or too high but is relatively uniform across interviewers, then the causes are general and are not traceable to specific interviewers.

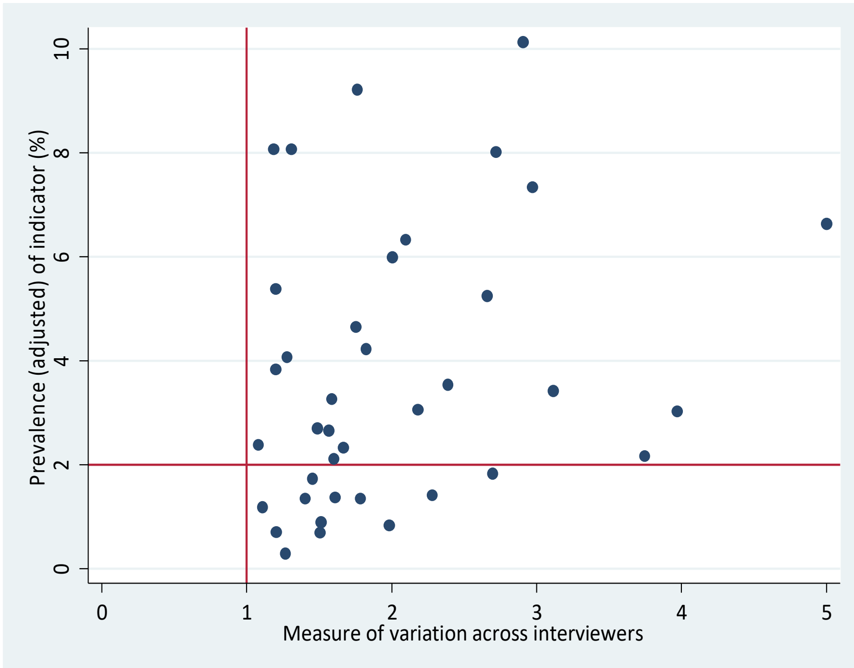
To determine whether the variation across interviewers is worth investigating, some standard is required. For this purpose we will use the level of significance of a test of the null hypothesis that the probability of the problematic outcome is the same for all interviewers, using the likelihood ratio chi-square statistic, in this context also known as the deviance D.

We will set a threshold for variation across interviewers at the 1% (or .01) level of significance. For most outcomes, this is relatively easy to reach just by chance, because of the large number of interviews. The number of interviewers (K) varies from one indicator to another and from one survey to another. The degrees of freedom (df=K-1) and the critical value also vary. In order to convey the magnitude of D we will divide it by the critical value of the chi-square distribution with the specified df and a .01 level of significance. If this chi-square ratio is greater than 1, then the variation across interviewers is statistically significant at the .01 level.

Figure 4.1 is a scatterplot in which each dot represents a combination of an indicator and a survey. There are five indicators and 15 surveys, but because the indicators of refusal are not relevant for all 15 surveys, the scatterplot contains only 38 dots, mostly referring to nonresponse rates. For each combination of indicator and country, we have two numerical values corresponding to the axes of the figure. The value shown on the vertical axis is the prevalence of the indicator, i.e., the percentage of relevant cases that are coded “1”. The value shown on the horizontal axis is the chi-square ratio described above.

Any chi-square ratios greater than 5—there is only one such value in this group of tests—is shown as 5 in the figure. (The point farthest to the right in the figure actually has a chi-square ratio of 10.22 rather than 5.) Points above the red horizontal line indicate a prevalence greater than 2%; points to the right of the red vertical line indicate variation across interviewers that is significant at the .01 level. All 38 dots are to the right of the red vertical line, implying that the variation across interviewers is always statistically significant at the .01 level.

Figure 4.1 Scatterplot of the prevalence of the five indicators of nonresponse and refusal (the vertical axis) and the measure of the significance of the variation across interviewers. The horizontal scale is the ratio of the calculated chi-square to the critical value for a .01 test of significance.



The lower right quadrant refers to 12 combinations of surveys and indicators for which the total deviance is significant at the .01 level but the prevalence is below the 2% threshold. These 12 combinations will not be identified explicitly. The 26 combinations represented by points in the upper right quadrant are presented in List 4.1, in the sequence of the indicators for this chapter and the country name (in alphabetical order). The two most extreme examples are in boldface type.

List 4.1 Combinations of country and indicator for which variation across interviewers is significant at the .01 level and the prevalence is above 2%. "Chi2 ratio" is the ratio of the calculated chi-square to the critical value for a .01 test of significance.

Country	Indicator	Prevalence (%)	Chi2 ratio
Armenia	Women nonresponse	2.2	3.7
Kenya	Women nonresponse	3.4	3.1
Liberia	Women nonresponse	2.7	1.6
Myanmar	Women nonresponse	4.2	1.8
Malawi	Women nonresponse	2.3	1.7
Timor-Leste	Women nonresponse	3.0	4.0
Tanzania	Women nonresponse	2.7	1.5
Uganda	Women nonresponse	3.0	2.2
Zimbabwe	Women nonresponse	3.8	1.2
Armenia	Men nonresponse	3.5	2.4
Burundi	Men nonresponse	2.4	1.1
Myanmar	Men nonresponse	9.2	1.8
Malawi	Men nonresponse	5.4	1.2
Nepal	Men nonresponse	4.1	1.3
Timor-Leste	Men nonresponse	5.2	2.7
Tanzania	Men nonresponse	8.1	1.2
Uganda	Men nonresponse	6.0	2.0
Zimbabwe	Men nonresponse	8.1	1.3
Zimbabwe	Women HIV test refusal	7.3	3.0
Burundi	Men HIV test refusal	2.1	1.6
Malawi	Men HIV test refusal	8.0	2.7
Zimbabwe	**Men HIV test refusal**	**10.1**	**2.9**
Armenia	**Women refusal of ht wt hb**	**6.6**	**10.2**
Myanmar	Women refusal of ht wt hb	4.6	1.8
Malawi	Women refusal of ht wt hb	3.3	1.6
Zimbabwe	Women refusal of ht wt hb	6.3	2.1

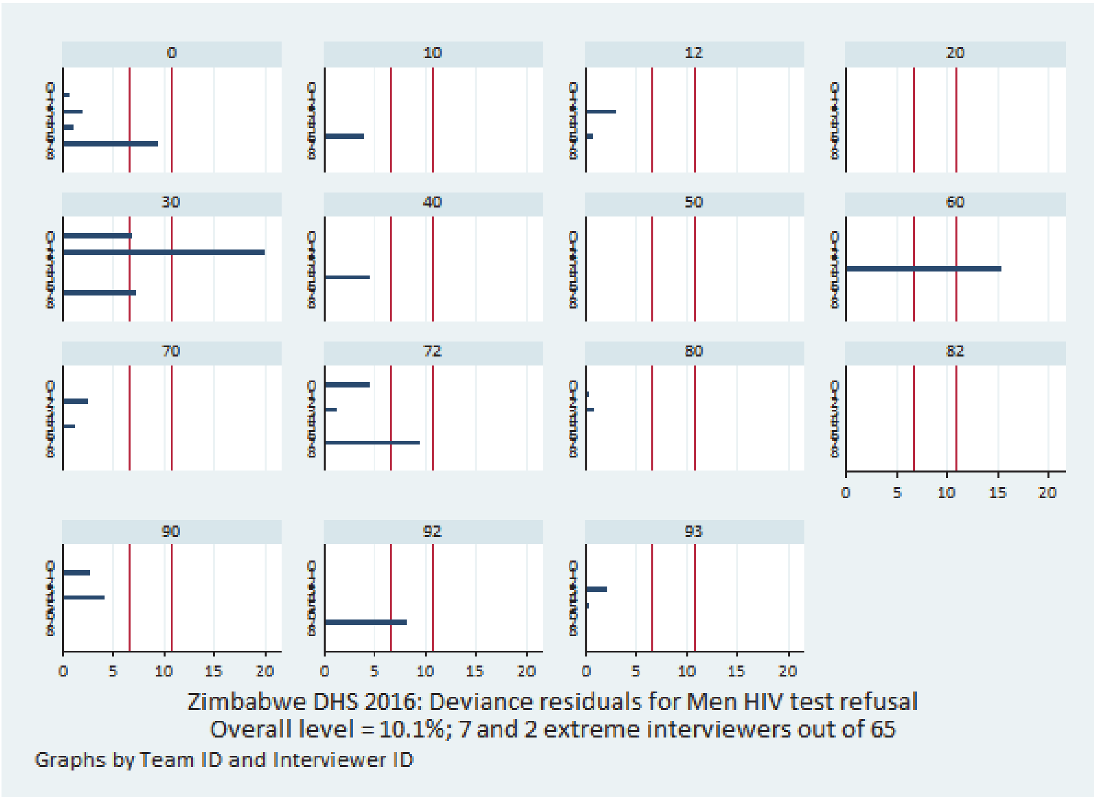
In most of the 26 problematic combinations, a relatively small number of interviewers appear to be responsible for much of the variation. We have prepared a figure for each combination to show visually the variation across interviewers. Because of space limitations, we will focus on just the two most serious examples of significant interviewer effects. In Figure 4.1, two points or combinations stand out—the point highest on the vertical axis and the point farthest to the right on the horizontal axis. These two combinations are examined further in Figures 4.2 and 4.3, respectively.

In Zimbabwe, the prevalence of men's HIV test refusal is 10.1%, and the chi-square statistic is 2.9 times as large as its critical value for a .01 test. Figure 4.2 shows the contributions to the total deviance that come from the 65 interviewers listed for this outcome. This figure is a set of horizontal bar graphs, with one sub-graph for each team and bars for the members of the teams. The team ID code that is given at the top of each subgraph has been inferred from the interviewer ID code.[7] The bars show the deviance residuals dr2 for the interviewers whose refusal rate was larger than the overall refusal rate. (As stated above, we do not flag the interviewers whose rate was less than the overall rate.) The horizontal axis ("x" axis) in each sub-graph is the magnitude of the deviance residual for that interviewer. Each sub-graph includes two vertical red lines. The first line marks the critical value of chi-square with one degree of freedom for a .01 test (6.63); the second line, to the right of the first one, marks the critical value for a .001 test (10.83). The caption "7 and 2 extreme interviewers out of 65" is intended to convey that seven horizontal bars cross the

[7] We do not know for certain that the interviewers in each subgraph actually worked together. For all of these surveys, the interviewer ID code has 3 or 4 digits, either abc or ab0c (in which the 10's digit is always zero). We infer that c is the code for an individual within team ab. It is possible that in some cases this unpacking procedure does not correctly identify the team number.

first red line and two of those also cross the second one. We see visually that those two bars go well beyond the .001 critical value. These interviewers had combinations of high prevalence and large number of cases.

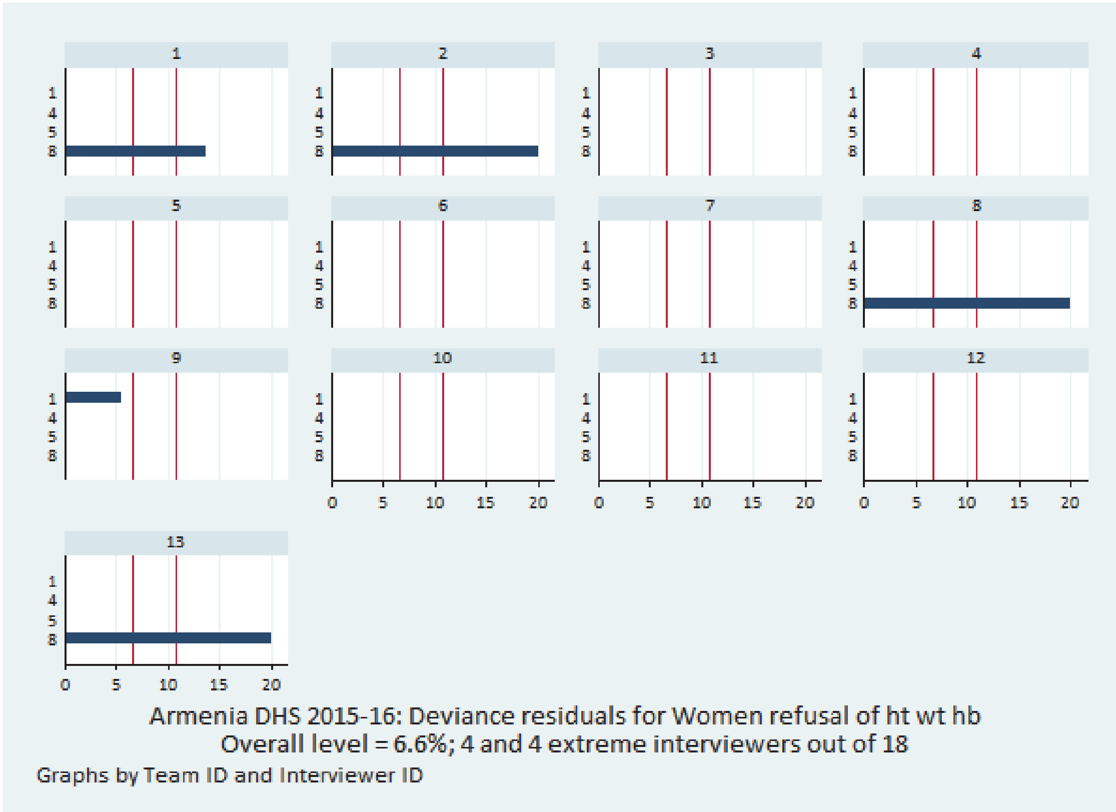
Figure 4.2 Deviance residuals dr2 for interviewers who are above the mean, for men's HIV test refusal in Zimbabwe, by interviewer within team.



As stated within the table, unique codes for 65 interviewers are specified in the household data file for this type of refusal, but clearly the figure does not include 65 bars. We repeat that bars are only shown for interviewers whose prevalence for this outcome was greater than the overall prevalence. In addition, deviance residuals close to 0 (dr2 is always positive) are not visible in the figure.

Another extreme combination of country and indicator, identified in List 4.1, is for women's refusal of ht wt hb, in Armenia. This refusal rate, 6.6%, is the highest of all the nine countries for which this specific indicator is available, and the chi-square statistic is more than 10 times its .01 critical value, the highest level of any combination of country and indicator in this chapter. Figure 4.3 shows the deviance residuals for the biomarker specialists for height, weight, and hemoglobin concentration in this survey. Among the 18 biomarker specialists, 4 stand out for extremely high levels of refusals. Three of them had deviance residuals in excess of 20 (the figures are truncated at 20). Deviance residuals have a chi-square distribution with one degree of freedom, and an expected value of 1 under the null hypothesis, so there is no question that these are highly significant deviations.

Figure 4.3 Deviance residuals dr2 for interviewers who are above the mean level of refusal for women's height, weight, and hemoglobin measurement in Armenia, by interviewer within team.

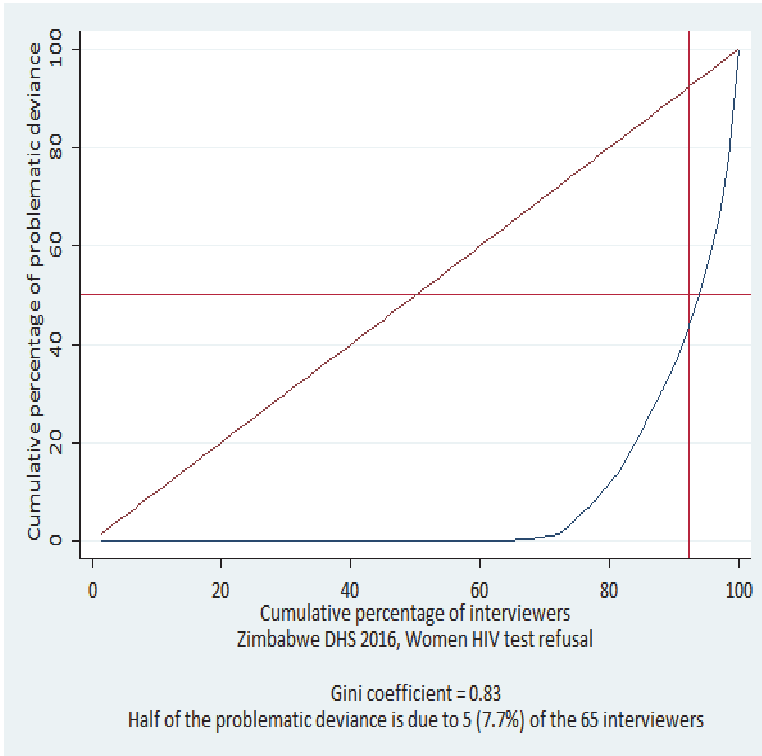


4.3 Concentration of Problematic Interviews

For nearly all combinations of surveys and indicators in List 4.1, the problems are concentrated in a small number of interviewers. We will assess this with two measures: the Gini coefficient of concentration and the percentage of interviewers who account for half of the problematic deviance—that is, who account for half of the deviance arising from interviewers whose prevalence of problematic responses is above the overall prevalence. Figures will be shown for the same two specific examples, as in Section 4.2.

Figure 4.4 refers to women's refusals for HIV testing in the Zimbabwe survey. Sixty-five interviewer ID codes are given for the variable indicating refusals. Out of those interviewers whose refusal rate was above the overall rates, five interviewers, or 7.7% of all the interviewers, accounted for half of the deviance. The Gini coefficient of concentration for the problematic deviance is 0.83. The figure provides a graphical interpretation of the concentration. The blue curve tracks the cumulative percentage of interviewers responsible for the cumulative problematic deviance, when the interviewers are ranked according to their respective contributions. Half of the problematic deviance is below the horizontal red line and the other half is above it. The vertical red line separates the interviewers into those responsible for the lower half of the problematic deviance from those responsible for the upper half; the intersection of the two red lines is approximately a point on the blue line.[8]

Figure 4.4 Concentration of problematic deviance for interviewers, for women's HIV test refusal in Zimbabwe.

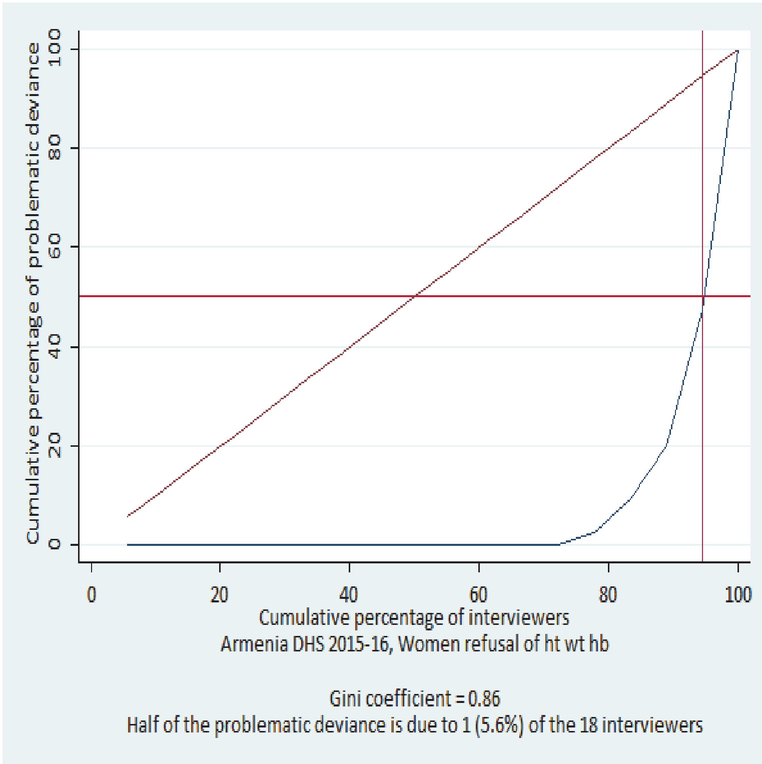


The Gini coefficient, a widely used measure of concentration, is the fraction of the triangular area below the diagonal line that is between the blue curve and the diagonal line. It ranges between 0 and 1.

Figure 4.5 refers to women's refusal of height, weight, and/or hemoglobin measurement in the Armenia survey. The measures of concentration of the problematic deviance are similar to those just discussed. The Gini coefficient is 0.86; half of the problematic deviance is due to a single interviewer, out of the 18 interviewers associated with this indicator.

[8] The three lines may not have a common point of intersection, simply because of the steepness of the blue line and the smoothing algorithm used by Stata.

Figure 4.5 Concentration of problematic deviance for interviewers, for refusal of women's height, weight, and hemoglobin measurement in Armenia.



It will be helpful to provide even more detail for the example of women's refusal of height, weight, and/or hemoglobin measurement in the Armenia survey, which appeared as the point farthest to the right in Figure 4.1 and was shown graphically in Figures 4.3 and 4.5. Because there were only 18 biomarker specialists, the diagnostic information can be shown compactly in List 4.2. Much of this report could be described as a synthesis of similar information for hundreds of such combinations of indicators and surveys.

List 4.2 **Interviewer-level distribution of women's refusals for height, weight, and/or hemoglobin measurement in the Armenia survey, with deviance residuals.**

Id	n1	n0	n	p1	dr	dr2	dr2pos
508	13	589	602	2.16	-5.09	25.94	0.00
708	12	542	554	2.17	-4.88	23.79	0.00
608	19	638	657	2.89	-4.31	18.57	0.00
908	21	650	671	3.13	-4.04	16.33	0.00
408	12	442	454	2.64	-3.86	14.90	0.00
1108	11	374	385	2.86	-3.33	11.12	0.00
1008	20	536	556	3.60	-3.13	9.82	0.00
1208	13	394	407	3.19	-3.08	9.50	0.00
505	0	2	2	0.00	-0.52	0.27	0.00
805	0	1	1	0.00	-0.37	0.14	0.00
1301	0	1	1	0.00	-0.37	0.14	0.00
804	0	1	1	0.00	-0.37	0.14	0.00
308	25	336	361	6.93	0.22	0.05	0.05
901	1	0	1	100.00	2.33	5.43	5.43
108	40	285	325	12.31	3.70	13.72	13.72
208	51	325	376	13.56	4.78	22.84	22.84
1308	86	444	530	16.23	7.59	57.61	57.61
808	93	311	404	23.02	10.56	111.43	111.43

In List 4.2, the interviewers are sorted by their deviance residual dr. The columns of the table are as follows:

id: interviewer identification code
n1: the number of responses for which the indicator is 1 (refusal)
n0: the number of responses for which the indicator is 0
n: the total number of responses for this indicator; n=n1+n0
p1: the percentage of responses for which the indicator is 1; p1=100*n1/n
dr: the deviance residual, before squaring
dr2: the deviance residual, after squaring
dr2pos: dr2 when dr>0; otherwise 0

The discussion of List 4.2 will be based on just a few specific observations, but the strategy could be extended broadly to the other indicators in the earlier List 4.1. Of the 18 interviewers, 14 had an ID code ending in the digit "8" and conducted hundreds of interviews; five did only one or two interviews and had an ID code ending in a digit other than "8". This is a common pattern across surveys, although the specific digits used for women or men or team leaders or biomarker specialists are not the same in every survey. In the Armenia survey, the final digit "8" consistently identifies the biomarker specialist on the team. In a handful of cases a non-specialist served as a substitute. One of these substitutes (#901) had only one entry and it was a refusal.

The dr (unsquared deviance residuals) values that are most negative are not at all problematic; these are interviewers who had very few refusals, in the range of 2.16% to 3.60%. It could be inferred that a "normal" level of refusals would be, say, 2% to 4%. dr and dr2 (the squared deviance residuals) are easily obtained from the generalized linear model described in Chapter 2 and Appendix 1.

The total deviance D for the model is 341.73, with 17 degrees of freedom. This is the sum of the dr2 terms in the next-to-last column of List 4.2. The critical value for a .01 LR chi-square test of the heterogeneity of interviewers is 33.41. The ratio of the observed value to the critical value is 10.22, which is the source of the "Chi2 ratio" of 10.2 shown in List 4.1 for Armenia women's refusal, the highest value in that list.

The high total deviance D is due to four interviewers (#108, #208, #1308, and #808), each of whom did hundreds of interviews with refusal rates in a range from 12% to 23%, contrasted with the "normal" range of 2% to 4%. These interviewers are the four identified with long horizontal bars in Figure 4.3. The worst case is interviewer #808, with a refusal rate of 23%. The sum of the dr2 values in the last column of List 4.2, which is limited to those value based on a positive dr, is 211.08. dr2 is 111.43 for interviewer #808, accounting for more than half of the 211.08. This is the interviewer ("1 of …18 interviewers") identified in the caption of Figure 4.5.

If we drop the most extreme interviewer (#808), and recalculate the coefficients, then P1=5.51%, representing a drop of a bit more than 1 percentage point in the overall level of refusals. The deviance declines to 217.58, and the chi-square ratio declines by a third, to 6.80. This ratio is still very large and would still be off the horizontal scale in Figure 4.1, but it is remarkable that a single interviewer could account for about a third of the total deviance and more than half of the problematic deviance.

If we drop all four of the extreme interviewers, we get an overall refusal rate of P1=3.16%, a little less than half of the original value of 6.63%. The deviance declines to D=25.06, with 13 degrees of freedom; the ratio of the deviance to its critical value is only 0.91. This would move the point in Figure 4.1 from being farthest to the right to being in the upper left quadrant, with a nonsignificant deviance, and much closer to the horizontal axis.

We do not believe it is possible to account for the high deviance of these four interviewers with the composition of their respondents. In terms of respondents' region, type of place of residence, wealth quintile, and other covariates, there are no obvious reasons why their refusal rates should be so high.

4.4 Associations with Interviewer Characteristics

Chapter 2 described the information collected about interviewers in the surveys that included a questionnaire for interviewers. We now investigate whether there is systematic variation in the prevalence of problematic interviews with respect to nonresponse and refusals, according to interviewer characteristics.

A complete analysis would examine the relationship between the covariates and the prevalence of every indicator of data quality in every survey. This report goes into much less detail. We pool all of the surveys, and also pool all of the indicators within a group of comparable indicators—in this chapter, all five indicators of nonresponse and refusals. The "collapsing" or aggregating of the data leads to a data structure that can be described for age, for example. Age of the interviewer has four categories, as described above. The data are first reduced to two numbers for each category of age. For example, for age interval 25-34, which is by far the interval with the most interviewers, "n" is the total number of times, in all surveys, and for all five items in this chapter, when an interviewer age 25-34 needed to record a respondent's eligibility, and n1 is the total number of times that the outcome was a nonresponse or refusal. We then calculate the percentage of nonresponse or refusal, P=100*n1/n. We also calculate the lower and upper ends of a 95% confidence interval (L and U, respectively). This is done within the glm framework described in Appendix 1, so there is also a chi-square test of the null hypothesis that there is no difference between the categories of a covariate. We do not include specific tests of the differences between pairs of categories, but such tests can be constructed from the confidence intervals.

We repeat that the Mozambique AIS survey is omitted from these results. For all covariates a category of "missing" is included; it is usually small. Figures 4.6a and 4.6b show P and the confidence intervals. Appendix Table A3.1 gives the numerical results. It is essential to keep in mind that the covariates are characteristics of the interviewers, not the respondents.

Figure 4.6a Graphical representation of the point estimates (P, a blue dot) and the lower and upper ends of 95% confidence intervals (L, a black dot, and U, a yellow dot, respectively) for the pooled percentages of all indicators of nonresponse and refusals, within categories of four interviewer characteristics: place of residence, age group, sex, and marital status. See Appendix Table A3.1 for numerical values.

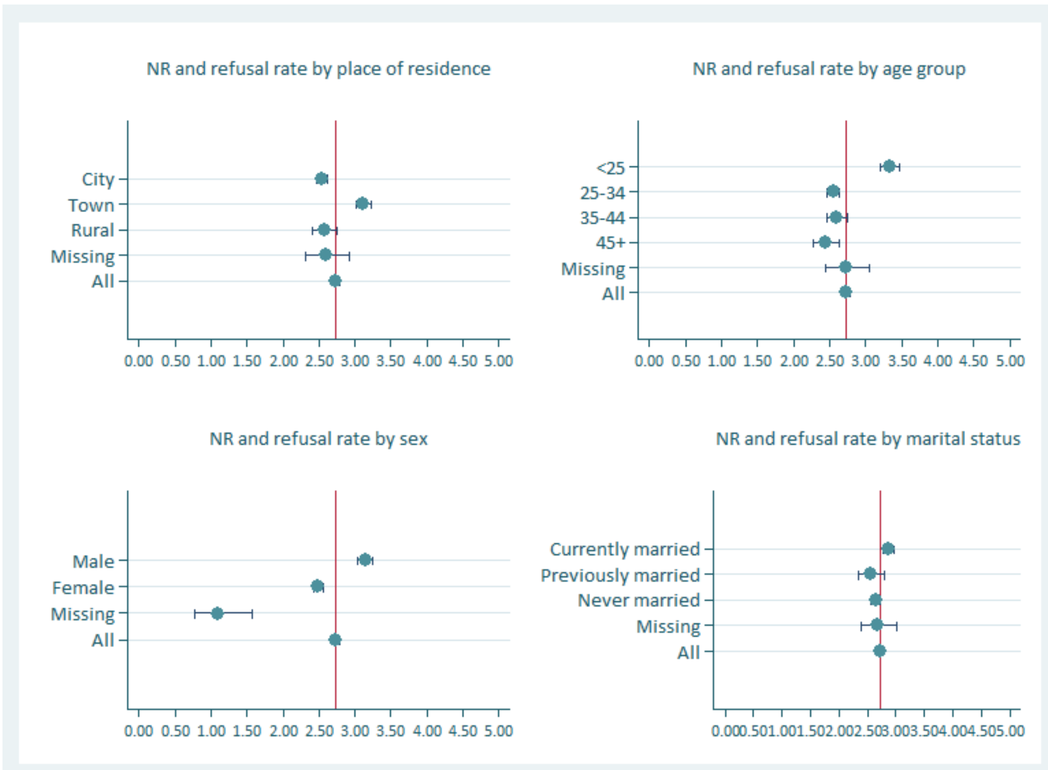
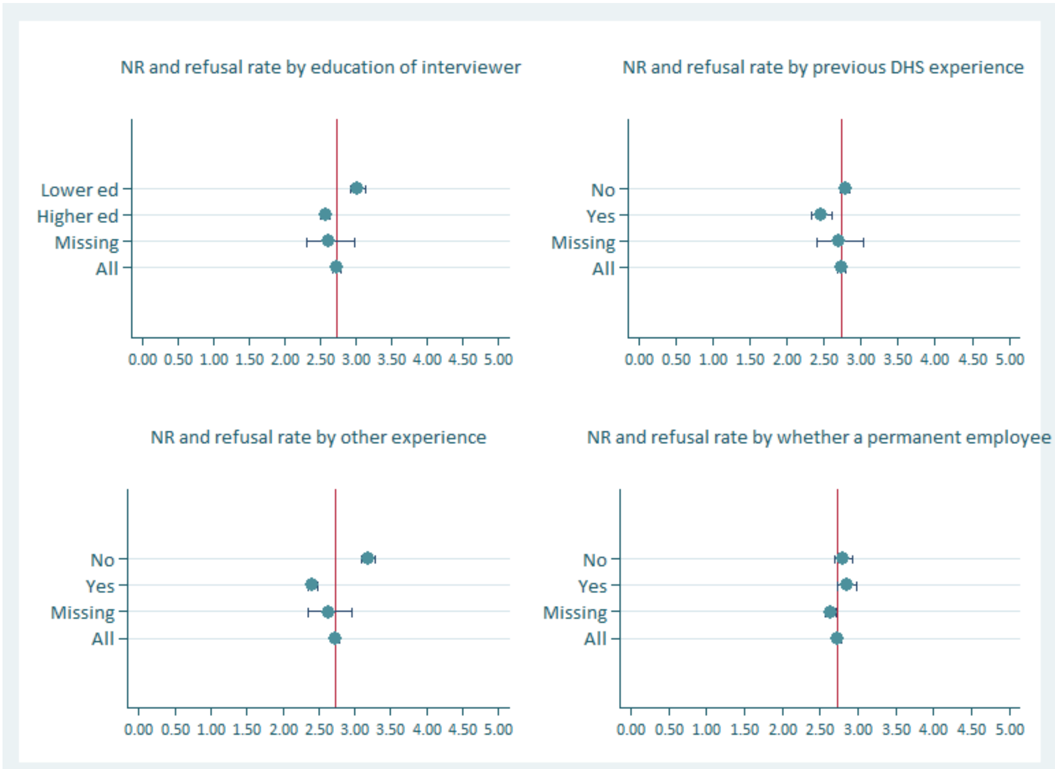


Figure 4.6b Graphical representation of the point estimates (P, a blue dot) and the lower and upper ends of 95% confidence intervals (L, a black dot, and U, a yellow dot, respectively) for the pooled percentages of all indicators of nonresponse and refusals, within categories of four interviewer characteristics: education, previous DHS experience, previous experience with other surveys, and whether a permanent employee of the implementing agency.



The overall level of nonresponse and refusals is 3.32%. Each subfigure within Figures 4.6 and 4.7 represents a separate glm model, for a separate combination of an outcome group (e.g., "NR and refusal rate") and covariate (e.g., "Place of residence") applied to a pooling of all the surveys. Each subfigure includes the estimate for "All", which is the same for all subfigures within a group, for easy reference. If a blue dot for a category of a covariate is to the right of the blue dot for "All", then the nonresponse in that category is higher than the overall level; if to the left, it is lower. Lower nonresponse is preferable, so the optimal category of a covariate will be the one farthest to the left. For the eight covariates in these figures, the patterns are easy to summarize:

Place of residence: Interviewers from rural areas, although they are a small minority of all interviewers, have significantly lower nonresponse and refusal rates. There is no difference between interviewers from cities and those from towns.

Age: There is a steady and monotonic reduction in the levels of nonresponse and refusals for older interviewers. The rate is highest for those under age 25.

Sex: The level of nonresponse and refusals is substantially lower for women than for men. This difference will be clarified in the next section of this chapter.

Marital status: There are no differences across categories of interviewer marital status. This is the only covariate which is not significantly related to the level of nonresponse and refusals.

Education: In this simple dichotomy, interviewers with a higher level of education have significantly lower levels of nonresponse and refusals than those with a lower level of education. This covariate is by far the best at discriminating between interviewers on this type of outcome. Because the distribution of this covariate is quite different in different countries, it may be capturing other unmeasured characteristics of those countries.

Previous DHS experience: Interviewers with DHS experience have significantly lower levels of nonresponse than interviewers without DHS experience.

Other previous experience: Interviewers with other survey experience have significantly lower levels of nonresponse than interviewers without such experience. General previous experience as an interviewer is more important than specifically DHS experience.

Permanent employee: Interviewers who are permanent employees of the implementing agency have slightly lower levels of nonresponse and refusals but the difference is not statistically significant. As a whole, this variable is statistically significant but only because of a large "missing" category, which has higher levels of nonresponse and refusals than either of the other categories. It is likely that "missing" tends to identify relatively inexperienced interviewers for whom "no" would be the appropriate category.

For the most part, these patterns are plausible and interpretable. Appendix Table A3.1 supplements the figures with numerical values. Except for the very small category of "missing" for sex of the interviewer, and the "less education" category, the prevalence of refusals and nonresponse always rounds to 2% to 4%. The single category with highest prevalence is "less education", which includes 44% of the interviewers and has a prevalence of 4.18%. The single category with lowest prevalence is "rural" place of residence, which applies to 10% of interviewers and for which the prevalence is 2.57%. Within this narrow range, all the covariates other than marital status are highly significant, with a p value less than .0001. The degrees of freedom for the chi-square test is the number of categories (including "missing") minus 1.

4.5 Sex Effects on Nonresponse for the Individual Interviews

The majority of DHS survey interviewers are women. Apart from biomarker specialists, who may be of either sex, the team may consist of, say, four women and one man. The sex composition of a team of interviewers is governed primarily (with a few exceptions) by the requirement that women must conduct individual interviews of women respondents, and men must conduct individual interviews of men. Household interviews can be conducted by either women or men, regardless of the sex of the household respondent—that is, the household member who provides the list of household members and information about the household as a unit.

The average length of time required for a household interview, a woman's interview, and a man's interview, will factor into the ideal balance of women and men on the team. The length of a household interview is affected by the number of household members. The length of an interview with a female respondent will be affected by such factors as her number of births, the number in the past 5 years, the number of episodes in the contraceptive calendar, and the number of siblings she has had if adult and maternal mortality rates are to be estimated. A man's interview is generally much shorter than a woman's interview, but will be lengthened if, say, a sibling history is required.

The typical pattern for assignments is as follows: After a team of interviewers arrives at a sample cluster or census enumeration area, and the households have been listed and the sample households selected, the interviewers are randomly assigned to the households. If the household interviewer is a woman, she will carry out the household interview and then proceed to interview the eligible women in the household. Similarly, if the household interviewer is a man, he will carry out the household interview and then proceed to interview the eligible men. It is then necessary for a male interviewer to complete interviews of men in the households that were first visited by a woman, and for a female interviewer to complete interviews of women in the households that were first visited by a man. Supervisors are responsible for ensuring that these subsequent reassignments according to the sex of the interviewer and the respondent actually take place.

There are four possible combinations by sex if an eligible woman or man is not, in fact, individually interviewed:

> The household interviewer is a woman; she is subsequently expected to interview a woman.
> The household interviewer is a woman; a male interviewer must be identified to interview a man.
> The household interviewer is a man; a female interviewer must be identified to interview a woman.
> The household interviewer is a man; he is subsequently expected to interview a man.

We can compare four nonresponse rates corresponding to these four combinations, in sequence:

> "Women NR" if the interviewer is a woman.
> "Men NR" if the interviewer is a woman.
> "Women NR" if the interviewer is a man.
> "Men NR" if the interviewer is a man.

There are potential combinations that we cannot differentiate, because if an individual interview is classified as "NR", we only have the ID code for the household interviewer. We cannot know with certainty that this is the interviewer who was supposed to do the individual interview but failed to do it. It is possible, for example, that a female interviewer, #1, does the household interview, but a different female interviewer, #2, is assigned to interview a woman in that household, simply because of workload or logistical considerations. This should not be typical, but it could happen. Then, insofar as the failure to interview the eligible female household member can be attributed to a specific interviewer, it could be that interviewer #2 was responsible for the nonresponse, rather than interviewer #1. However, we do not know the ID for interviewer #2. We only know the ID for interviewer #1. Similarly, the failure to interview an eligible male could be due to a second but unidentifiable male interviewer.

Figure 4.7 is a scatterplot showing the unsquared residual deviances (dr) for nonresponse of women and men when the household interviewer is a woman. The vertical axis is the deviance residual for men and the horizontal axis is the deviance residual for women. This choice of axes treats the women as the reference category, because in all surveys the principal interest is in the women's interviews. The diagonal line represents equality of the two nonresponse rates. If a point is above the diagonal line, then the nonresponse rate was higher for men than for women, and if the point is below the diagonal line, it was lower. Vertical and horizontal red lines are drawn at the critical values for a two-tailed .01 test of significance, +/-2.58. The great majority of interviewers have dr values that are inside these limits on both axes, but there are clearly more points above the diagonal than below it. Of the 642 female household interviewers in this study, 396, or 62%, had a higher nonresponse rate for men's interviews than for women's interviews. The excess of points above the diagonal line is statistically very significant.

Figure 4.7 Deviance residuals (dr) for nonresponse for the women's and men's interviews, for all female household interviewers

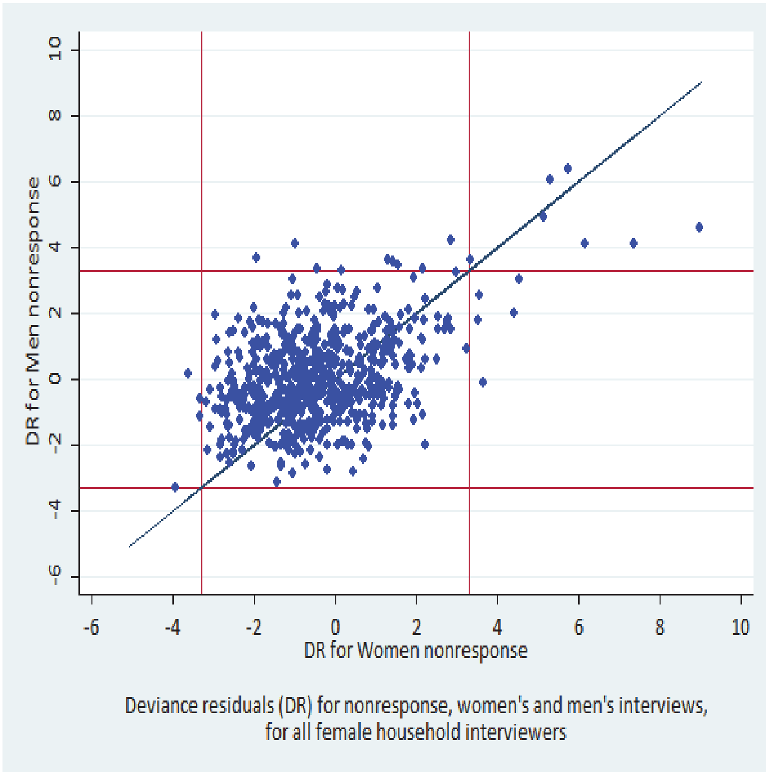
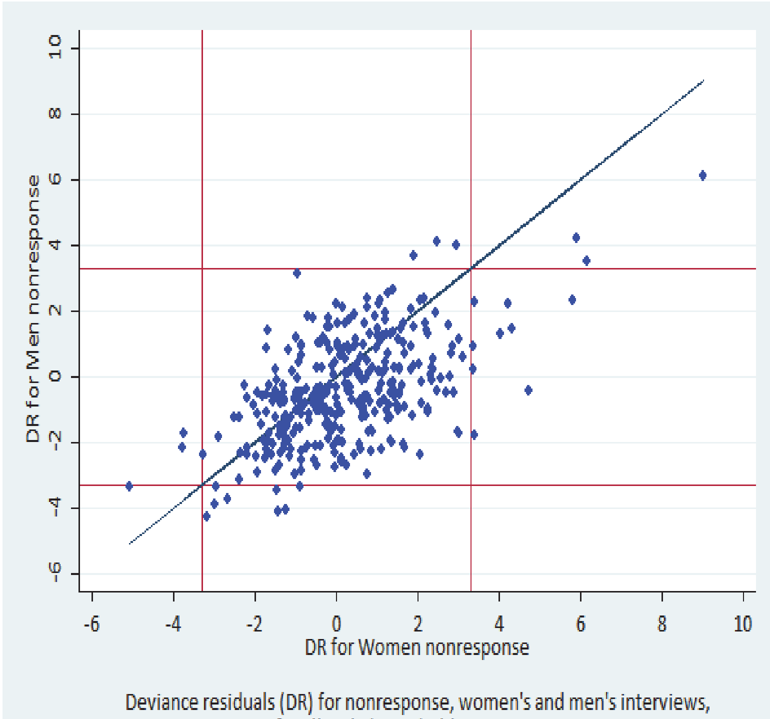


Figure 4.8 shows the corresponding scatterplot for male household interviewers. A majority of the points are below the diagonal line, the reverse of the pattern in Figure 4.8. Of the 337 male interviewers, 227, or 67%, are below the line, implying that the nonresponse rate among respondents eligible for the individual interview tended to be lower for interviews of men than for interviews of women. Because the household interviewers are primarily women, the overall nonresponse rate for eligible men is generally higher than that for eligible women. The excess of points below the diagonal line is statistically very significant.

This analysis could be extended to examine whether the household respondent is the same sex as the interviewer and is, or is not, an eligible respondent. Whatever the reason, it appears that the response rate is better when the household interviewer is the same sex as the eligible respondent identified in the household interview.

Figure 4.8 Deviance residuals for nonresponse, women's and men's interviews, for all male household interviewers



Deviance residuals (DR) for nonresponse, women's and men's interviews, for all male household interviewers

5 UNDER-5 MORTALITY

The birth history obtained during the interview with eligible respondents includes information about child survival. If a child died, the mother is asked the child's age at death. The responses are given in days up to 30 days; in months up to 23 months; and then in years. The responses are interpreted as completed age (in days, months, or years). DHS main reports include estimates of five mortality rates for children under age 0-4, based on exposure to risk and deaths observed during the 5 years (60 months) before the interview. The rates refer to the following five age intervals: 0 months (neonatal mortality); 1-11 months (post neonatal mortality); year 1, or 0-11 months (infant mortality); years 1-4 (child mortality); and years 0-4 (under-5 mortality). The calculation of the rates is complex and is described in the Guide to DHS Statistics (2018).

The main measurement issues related to child mortality are inaccurate reporting of age at death and omission of a death. Omission is more serious. If a child who died is omitted, then the birth is also omitted, and fertility rates as well as mortality rates will be underestimated. Omission is believed to be more likely for children who died young.

5.1 Indicators

We consider just three indicators, which will only be calculated for the surveys that had complete birth histories, and therefore this chapter omits the AIS and MIS surveys. All of the interviewers associated with this outcome are women, except for a handful of instances in which a man interviewed a woman respondent. In the absence of misreporting, the true means of these indicators would not be zero.

One indicator of displacement of age at death

Heaping on death at 12 months of age: died_12_months. In the birth history, age at death is coded for children who died. Age at death is to be given as *completed* days, months, or years. Thus, the first 24 hours is day 0, the first month is month 0, and the first year is year 0. The responses are expected to be in days, up to 30 days; then in months, up to 23 months; and then in years of age. There is often a concentration at 12 months, which is interpreted as partially due to ambiguity about whether the response is actually being given in completed months, or is a translation of "one year" into 12 months. Over-reporting at 12 months, for deaths that actually occurred at 11 months, will spuriously deflate the infant mortality rate and inflate the rate for age 1-4, but will not affect the under-5 mortality rate. The numerical effect is small because the number of deaths reported at exactly 12 months is generally small relative to the number in months 0-11 or in months 12-59. The indicator is 1 if a death was reported at exactly 12 months of age. It is 0 for all other deaths that were reported in days or months. It is NA for children who survived or whose death was reported in years. In countries with low child mortality, such as Armenia, this statistic is unstable, especially when disaggregated.

Two indicators of omission of deaths

Death in the first month: died_month_0. This is the first of two indicators of potential omission of children who died. Interviewers who deviate from the general pattern of reporting in a survey, particularly by identifying (relatively) fewer child deaths, may have inadequately probed for child deaths, i.e., for the identification of live births, in the birth histories, that resulted in a child death. Such omissions are difficult to detect (see, for example, Pullum and Becker 2014) but are important because they spuriously reduce the death rates for children and (to a lesser extent) fertility rates. It is generally believed that omission is more serious for children who died young, especially in the first month. The indicator is 1 if a death was reported at exactly 0 months of age. It is 0 for all other deaths reported at age 1-11 months. It is NA for children who survived or whose death was reported at a completed age greater than 0 years. There is no restriction on the reference interval of time; we include all deaths in the birth history.

Death in the first year: died_year_0. The indicator is 1 if a death was reported at age 0-11 months. It is 0 for all other deaths that were reported at age 12-59 months. It is NA for children who survived or whose death was reported at a completed age greater than 4 years. There is no restriction on the reference interval of time and there is no adjustment for censoring, that is, for the reduced exposure to the risk of death for the most recent births.

The indicators of omission do not describe death rates, i.e., ratios of deaths to births; rather, they describe the age distribution of deaths, and are less affected by the underlying level of child mortality than the death rates would be. It is known that the age distribution of deaths tends to vary somewhat according to the underlying level of mortality (see, for example, Hill and Choi 2006), especially with neonatal deaths becoming a greater share of infant deaths as infant deaths decline, and infant deaths becoming a greater share of under-5 deaths as under-5 deaths decline, but that relationship will not be taken into account.

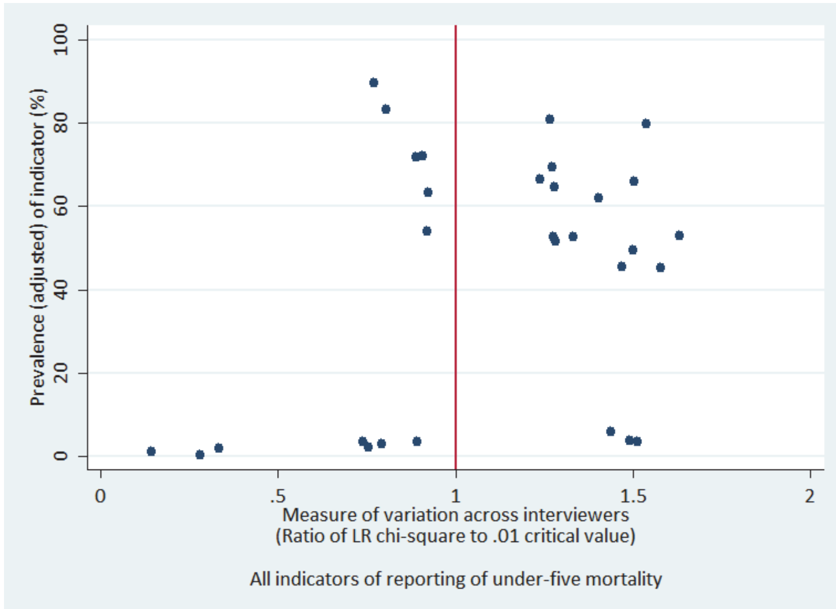
Because the two indicators that identify omission have a value of "1" for cases where we believe omission is more likely, there is a risk of reversing the interpretation—for the two indicators of omission, and for these alone within this report, *a higher percentage of responses with "1" is desirable*. That is, under the assumption that omission of deaths is primarily the omission of *early* deaths, we become suspicious of omission if the interviewer reports few early deaths. When this percentage is low, there is evidence of omission; when it is high, there is evidence of more complete reporting of child deaths.

5.2 Identifying Problematic Surveys and Indicators

The indicators related to under-5 mortality can—and should—have values greater than zero even in the absence of any reporting errors. The effect of errors will tend to be in different directions. For the first indicator, the relative frequency of responses at age 12 months for age at death, error is usually described as an excess of such responses. In this report, as with all the heaping indicators, both over-reporting and under-reporting will be considered problematic. For the other two indicators, using the age distribution of deaths to identify potential omission, only low values will be identified as problematic. For all three indicators, a high level of dispersion across interviewers within a specific survey suggests reporting errors.

Figure 5.1 is a scatterplot showing the levels of all three indicators on the vertical axis and the degree of dispersion on the horizontal axis. As before, dispersion is measured by the ratio of the chi-square statistic, for a logit or glm regression, to its critical value for a .01 test. Dots to the right of the red vertical line indicate interviewer effects that are significant at the .01 level. Because the indicators do not have natural zeroes, there is no horizontal line for a threshold, such as 2%, beyond which the prevalence is clearly too high to be acceptable. The dots in the lower tier refer to the indicator of heaping at age 12 months; only 3 out of 10 are statistically significant, partly because of the small numbers of cases. The dots in the upper tier refer to the other two indicators; 14 out of 20 are significant.

Figure 5.1 Combinations of country and indicator for which variation across interviewers is significant at the .01 level. The horizontal scale is the ratio of the calculated chi-square to the critical value for a .01 test of significance



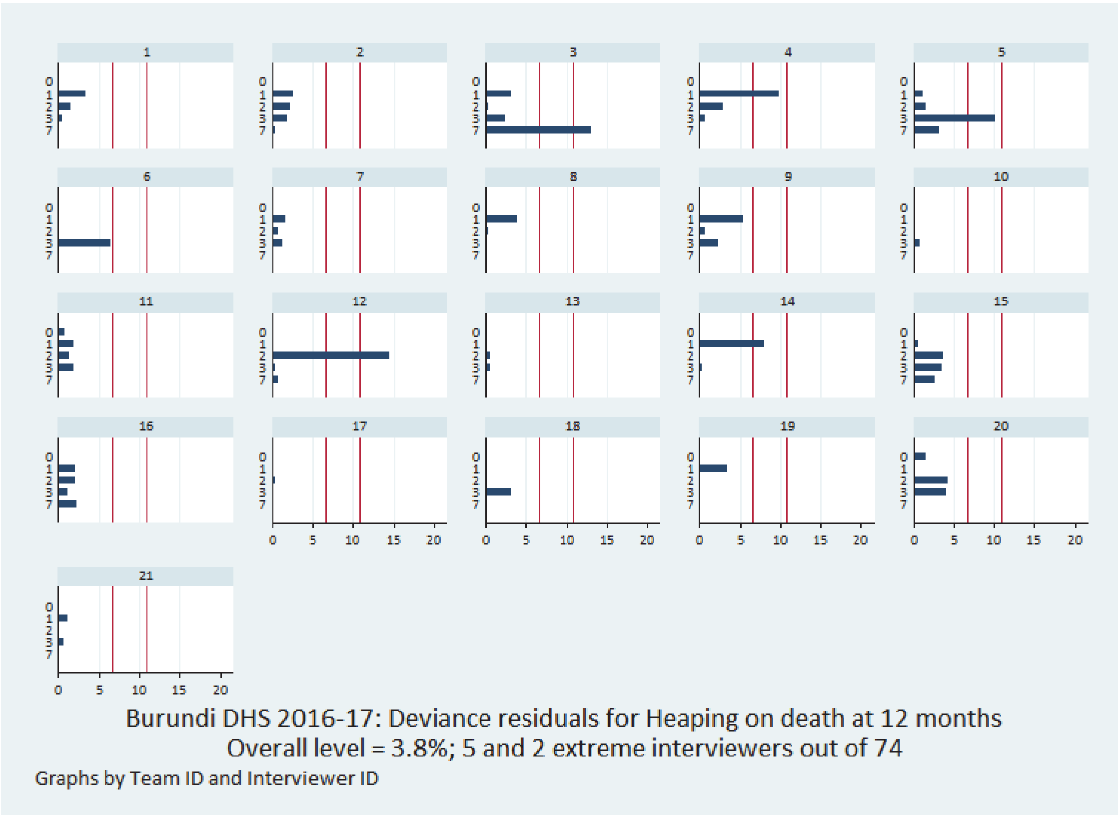
List 5.1 Combinations of country and indicator for which variation across interviewers is significant at the .01 level. "Chi2 ratio" is the ratio of the calculated chi-square to the critical value for a .01 test of significance

Country	Indicator	Prevalence	Chi2 ratio
Burundi	**Heaping on death at 12 months**	**3.8**	**1.5**
Malawi	Heaping on death at 12 months	5.8	1.4
Uganda	Heaping on death at 12 months	3.6	1.5
Burundi	Death in first month	45.4	1.5
Myanmar	Death in first month	52.7	1.3
Malawi	Death in first month	51.5	1.3
Nepal	Death in first month	66.5	1.2
Timor-Leste	**Death in first month**	**52.9**	**1.6**
Tanzania	Death in first month	49.4	1.5
Uganda	Death in first month	45.2	1.6
Zimbabwe	Death in first month	52.7	1.3
Burundi	Death in first year	64.6	1.3
Myanmar	Death in first year	79.6	1.5
Malawi	Death in first year	61.9	1.4
Timor-Leste	Death in first year	80.6	1.3
Tanzania	Death in first year	69.3	1.3
Uganda	Death in first year	65.8	1.5

List 5.1 identifies the 17 combinations of country and indicator that are to the right of the red vertical line in Figure 5.1, which indicates statistical significance at the .01 level. All three indicators appear, but the indicator associated with neonatal mortality—deaths in the first month—dominates. Two combinations are bolded because of their high values of the Chi2 ratio: deaths in month 12 in Burundi and deaths in month 0 in Timor-Leste. We will only provide more detail for these two combinations, although everything on List 5.1 that is statistically significant could be explored further. Note that Burundi also has significant dispersion for deaths in the first month and deaths in the first year, and Timor-Leste also appears for deaths in the first year. Figures 5.2 and 5.3 provide graphical representations of the contributions to the total

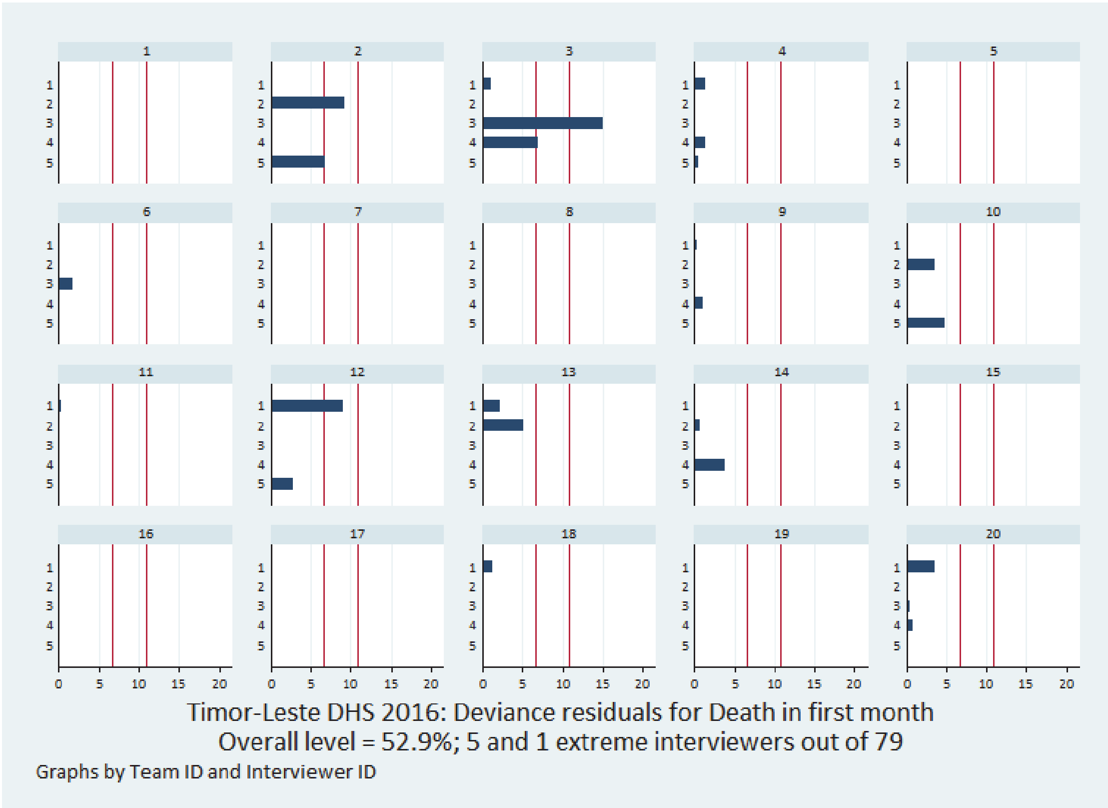
deviance for the respective indicators, in the Burundi and Timor-Leste surveys, that come from specific interviewers.

Figure 5.2 Deviance residuals dr2 for interviewers, for heaping of deaths at 12 months in Burundi, regardless of whether there is over-reporting or under-reporting, by interviewer within team



In Figure 5.2, the total deviance in the first indicator in Burundi is allocated across the interviewers who collected birth histories that included deaths. Five of them have squared deviance residuals dr2 that exceed the .01 level of significance, and two interviewers—#57 and #122—exceed the .001 level. Figure 5.2 identifies the interviewers in Timor-Leste whose proportion of first-year deaths that were also first-month deaths is significantly less than the overall proportion. Five interviewers were significantly low at the .01 level and one of them, #33, was significantly low at the .001 level.

Figure 5.3 Deviance residuals dr2 for interviewers who are below the mean, for potential under-reporting of deaths in the first month in Timor-Leste, by interviewer within team



5.3 Concentrations of Problematic Interviewers

This section further describes the concentration of potential problems with the reporting of child deaths, using the Gini coefficient of concentration and the percentage of interviewers who account for half of the problematic deviance. Figures 5.4 and 5.5 continue with the two most extreme examples identified in Figure 5.1. Figure 5.4 refers to deaths in month 12 in the Burundi survey. The Gini coefficient for the distribution of the total deviance is extremely high, at 0.88, and just three interviewers accounted for half of this. Figure 5.5 refers to Timor-Leste. The concentration is similar to that in Burundi. The Gini coefficient is 0.85, and five of the interviewers accounted for half of the problematic deviance.

Figure 5.4 Concentration of problematic deviance for interviewers, for deaths in month 12 in the Burundi survey

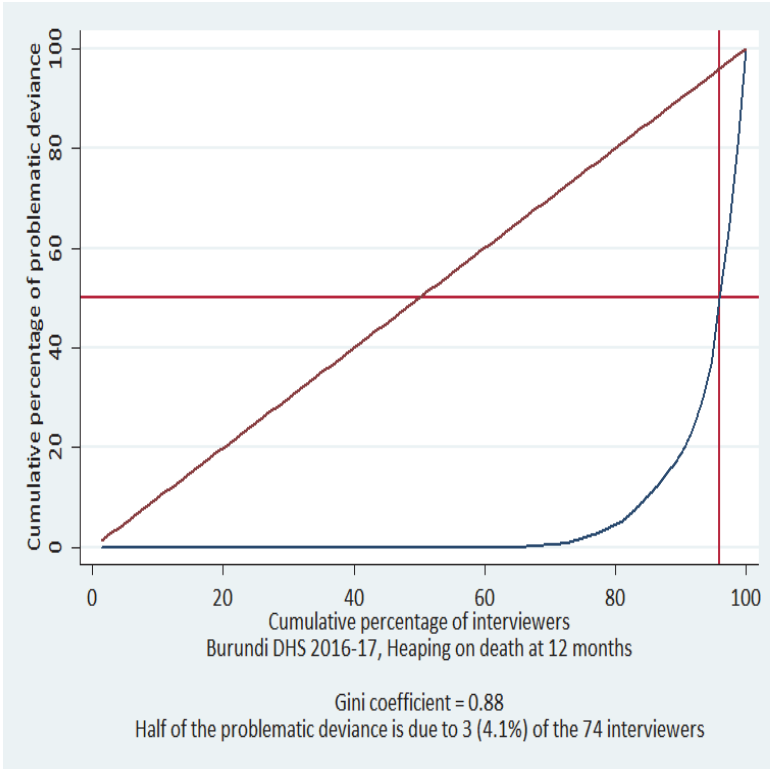
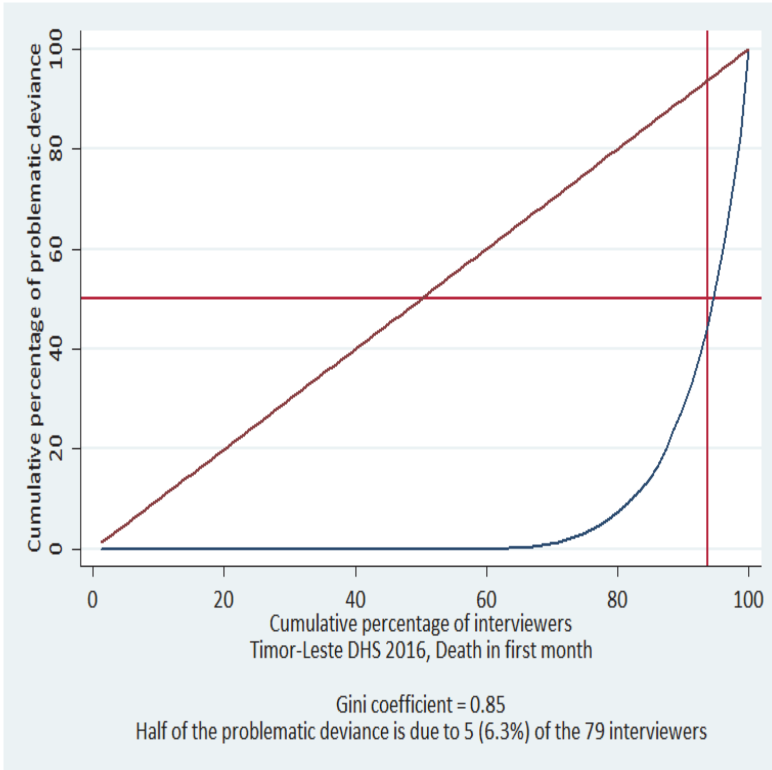


Figure 5.5 Concentration of problematic deviance for interviewers, for deaths in month 0 in the Timor-Leste survey



Because potential omission of children who died is a particular concern for DHS surveys, List 5.2 provides more detail on the indicator for potential under-reporting of deaths in month 0 for individual interviewers in the Timor-Leste survey, which is identified in Figure 5.1 as the point farthest to the right. The list omits the interviewers for whom the deviance residual was in the range of -1.96 to +1.96.

In the Timor-Leste survey, the overall prevalence—that is, the percentage of deaths in the first year that were also reported to be in the first month—is 52.9%. The sign of the deviance residual dr will be negative if the prevalence for an individual interviewer is less than 52.9%, and the magnitude of dr will increase if the deviation below 52.9% is wider and the number of cases is larger. The five interviewers with the largest negative deviations are #33, #22, #121, #34, and #34. These interviewers are also identifiable in Figure 5.3. Interviewer #33 is conspicuous because, out of 15 deaths reported in the first year, only 1 was in the first month. Interviewer #121, who is third on the list, had a much higher prevalence, p1=28.95%, but the negative deviation has a substantial impact on the deviance because it is based on a larger denominator, n=38 deaths in the first year, only n1=11 of which were assigned to the first month. Low prevalence (p1) may indeed identify omission of deaths in the first year, but the number of cases per interviewer, n, is numerically too small to yield convincing evidence of omission by the specific interviewers with the lowest prevalence.

If the five interviewers with the largest negative values of dr are dropped, then the percentage of infant deaths that are neonatal rises from 52.9% to 56.0%, and the Chi2 ratio declines from 1.63 to 1.04, meaning that the total deviance becomes only marginally significant at the .01 level.

A few interviewers are at the opposite end, or bottom of List 5.2—those with relatively high proportions of deaths in month 0. These interviewers may well be problematic too. The most extreme five interviewers in a positive direction (#12, #103, #154, #174, #71) recorded a total of 54 infant deaths, of which 48 were reported to be neonatal deaths. This pattern is also suspicious; we would have expected about 29 of them to be neonatal. A possible interpretation that encompasses both ends of List 5.2 is simply that some interviewers tend to misreport age at death, rather than to omit deaths. In the aggregate, interviewers who under-report the first month among infant deaths are offset by other interviewers who over-report the first month.

List 5.2 Interviewer-level distribution of reported deaths in month 0 in the Timor-Leste survey, with deviance residuals. Interviewers with dr in the range -1.96 to +1.96 are omitted

id	n1	n0	n	p1	dr	dr2	dr2neg
33	1	14	15	6.67	-3.87	15.01	15.01
22	4	16	20	20.00	-3.03	9.17	9.17
121	11	27	38	28.95	-2.99	8.94	8.94
34	1	8	9	11.11	-2.65	7.04	7.04
25	2	10	12	16.67	-2.61	6.79	6.79
132	4	12	16	25.00	-2.27	5.17	5.17
105	3	10	13	23.08	-2.20	4.83	4.83
…							
155	15	5	20	75.00	2.03	4.14	0.00
143	30	13	43	69.77	2.25	5.08	0.00
95	15	4	19	78.95	2.36	5.57	0.00
71	5	0	5	100.00	2.52	6.37	0.00
174	5	0	5	100.00	2.52	6.37	0.00
154	19	5	24	79.17	2.68	7.16	0.00
103	10	1	11	90.91	2.75	7.54	0.00
12	9	0	9	100.00	3.39	11.46	0.00

5.4 Associations with Interviewer Characteristics

Like other problems such as nonresponse, refusals, and displacement out of eligibility, omissions of child deaths probably occur partly from an interviewer's motivation to reduce the workload. They occur more readily when the respondent is not confident about a response or is reluctant to provide information, but the impact of the respondent's uncertainty will be amplified if the interviewer does not adequately probe, appears uninterested or rushed, etc. Thus, it would be useful to understand how the probability of omission is related to the characteristics of the interviewers.

When early deaths are a relatively small percentage of all deaths reported for age 0 or age 0-4, we cannot be sure whether this is a symptom of omission or of incorrect classification by age at death, or of some other effect that is less directly related to data quality, such as the general level of child mortality in the subpopulation being interviewed. However, we will proceed with the two indicators that compare deaths in the first month of age with deaths in the first year, and deaths in the first year with deaths in the first 5 years, respectively, to investigate whether they are related to the measured characteristics of the interviewers. The two indicators have a positive correlation of 0.20, which is not high but is significant at the .0001 level. A positive correlation would be expected if there is a tendency to omit early deaths, within either age interval. A positive correlation would also be expected if there is a tendency to displace events in the same direction, within either age interval, although there would seem to be less incentive or opportunity to displace than to omit. It could also result from the general relationship between the level and timing of child mortality.

We investigate the association between the covariates and the reported timing of deaths by pooling the two indicators of timing and applying glm models, as described in Appendix 2. The analysis includes two covariates that could potentially be related specifically to the collection of data on complete birth history—whether the interviewer has any living children and whether the interviewer has ever had a child that died. Figures 5.6, 5.7, and 5.8 present the results, with numerical values in Appendix Table A3.2.

Figure 5.6a Graphical representation of the point estimates (P, a blue dot) and the lower and upper ends of 95% confidence intervals (L, a black dot, and U, a yellow dot, respectively) for the pooled percentages of infant deaths that are also neonatal deaths, and under-5 deaths that are also infant deaths, within categories of four interviewer characteristics: place of residence, age group, sex, and marital status. All surveys are pooled.

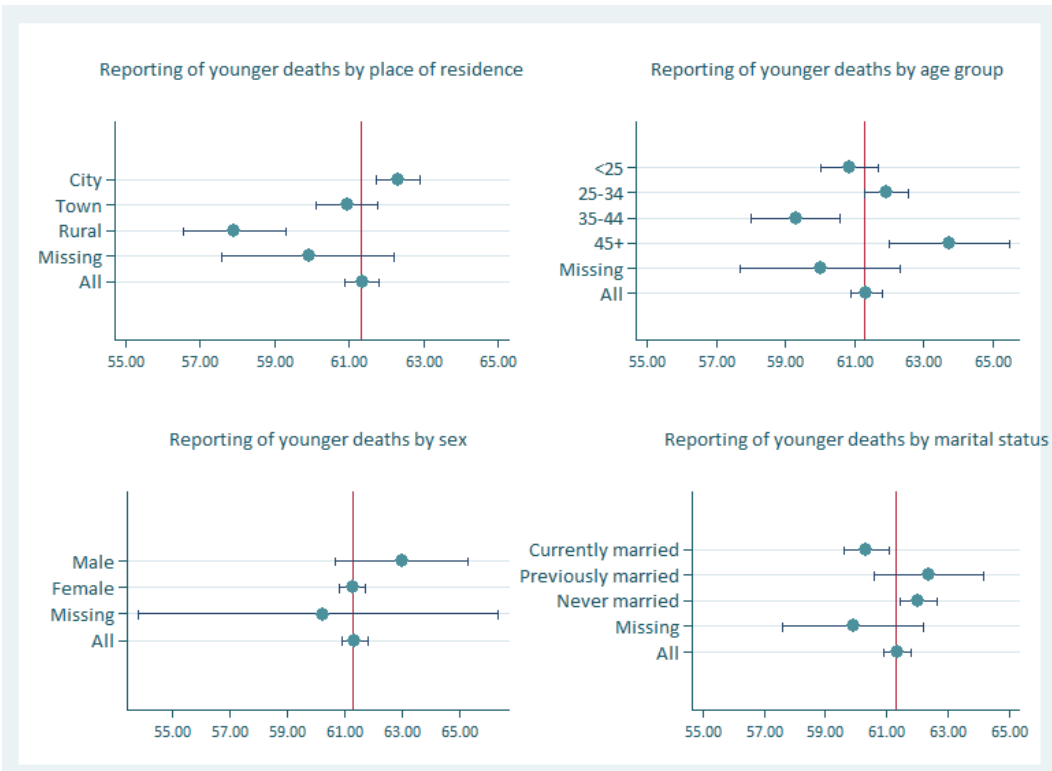


Figure 5.6b Graphical representation of the point estimates (P, a blue dot) and the lower and upper ends of 95% confidence intervals (L, a black dot, and U, a yellow dot, respectively) for the pooled percentages of infant deaths that are also neonatal deaths, and under-5 deaths that are also infant deaths, within categories of four interviewer characteristics: education, previous DHS experience, previous experience with other surveys, and whether a permanent employee of the implementing agency. All surveys are pooled.

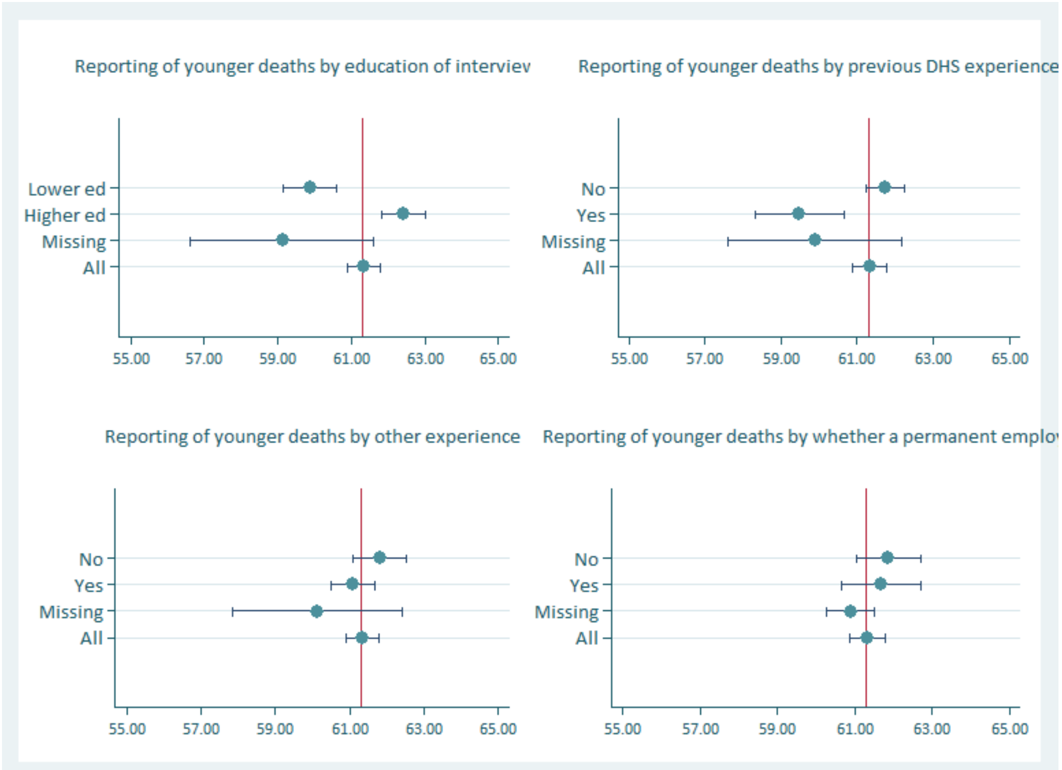
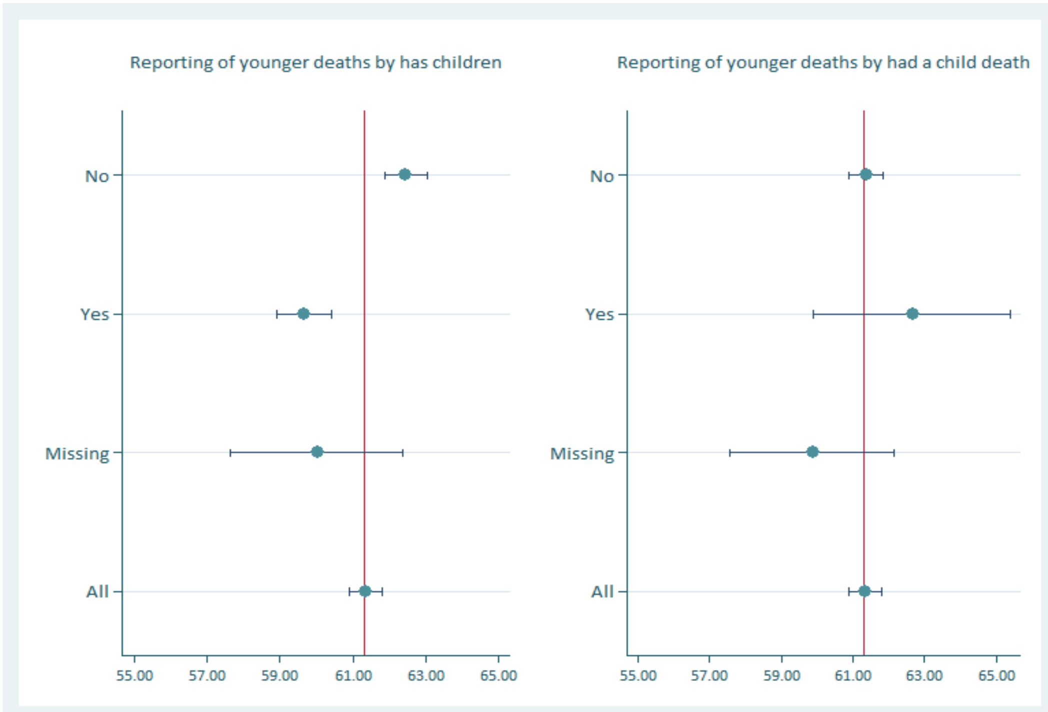


Figure 5.6c Graphical representation of the point estimates (P, a blue dot) and the lower and upper ends of 95% confidence intervals (L, a black dot, and U, a yellow dot, respectively) for the pooled percentages of infant deaths that are also neonatal deaths, and under-5 deaths that are also infant deaths, within categories of four interviewer characteristics: whether the interviewer has any children and whether the interviewer has experienced a child death. All surveys are pooled.



Each subplot in Figure 5.6a, 5.6b, and 5.6c is produced by a separate model that includes a categorical covariate. Each covariate includes a category for missing values, which produces wide confidence intervals because of the small numbers of missing cases. Any categories of covariates that have fewer cases will have wider confidence intervals. The subplots also include a row for the overall percentage, which is 61.34%, and a red vertical line through the mean, for easy visual reference. This is the overall pooled percentage of deaths that are reported to be early in the first year and/or early in the first 5 years. If a category has a lower percentage, we infer that interviewers in that category have a higher probability of omission (although we have suggested above that the percentage may vary for other reasons). In other words, categories showing prevalence above the mean are interpreted as having higher-quality data. As mentioned, for these indicators, and these alone within this report, a higher percentage is desirable. A lower level suggests omission of deaths.

Six of the ten covariates have a significant relationship with the outcome. When significance is found, it is at a very high level, .002 or better. The effects of the interviewer characteristics are as follows:

Place of residence: The prevalence is highest for “city” residents, lower for “town” residents, and lowest for “rural’ residents, and all of these differences are significant. We infer that urban interviewers tend to obtain more complete information about child deaths compared with rural interviewers.

Age: The effect of age is highly significant, but the pattern across age groups is uneven and difficult to interpret. The prevalence is lowest for age 35-44 and highest for age 45+. The younger age intervals are intermediate. We infer that interviewers age 45+ tend to obtain the most complete information about child deaths—significantly more complete than the interviewers under age 18 or age 35-44—but only about 7% of the cases are in this age interval.

Sex: Not significant. Only 3.7% of the cases (birth histories with any deaths under age 5) had male interviewers. The covariate is included to call attention to the fact that in some countries it is possible for men to interview women.

Marital status: The prevalence is lowest for "currently married" and highest for "never married". The "previously married" category of interviewers is small and cannot be distinguished from either of the other categories.

Education: Interviewers who have more education have higher prevalence than those with less education, by a substantial margin. The inference is that interviewers with better education tend to collect more complete information about child deaths.

Previous DHS experience: The interviewers with no DHS experience have a significantly higher percentage of reporting of early deaths, compared with the 15% of interviewers with DHS experience. The difference is approximately 2%. The inference is that previous DHS experience is not an advantage in obtaining complete information about child deaths.

Other survey experience: Not significant. Interviewers with DHS experience tend also to have other survey experience.

Permanent employee: Not significant.

Has children: Interviewers who do not have living children have a significantly higher prevalence than those who do not, and the difference is substantial, at about 3%. Because the percentage of interviewers who have had a child death is small, this covariate is essentially equivalent to whether the interviewer has ever had a child. It could have been expected that the interviewers who have children would be able to establish better rapport with respondents during the collection of the birth histories, and therefore obtain more complete birth histories, but we do not find support for such an effect.

Had a child death: Not significant. The percentage of interviewers who reported having experienced a child death is small, at less than 3%. A characteristic with such a skewed distribution is unlikely to be useful for any statistical purposes.

To summarize these findings, interviewers who are urban, older, and more educated report significantly higher percentages of early deaths. It is plausible that such interviewers will tend to be better at collecting birth histories. However, the fact that interviewers who have no children, and interviewers who have no previous DHS experience, also have higher percentages, is harder to interpret.

6 REPORTING OF AGES AND DATES

Accurate reports of ages and dates are crucial for most DHS indicators. DHS has prepared several studies of the quality of age and date reporting. One of these (Pullum and Staveteig 2017) provides considerable detail on the interview strategies for obtaining ages and dates, on the interpretation of incompleteness codes, and on the imputation procedures for incomplete dates. Those details will not be repeated here.

A reason for including so many indicators of potential problems in the reporting of ages and dates is to identify the specific indicators within the different types that may be most useful for diagnosing data quality and interviewer effects. We also investigate whether the different types of indicators tend to be associated or to be independent of one another.

6.1 Indicators

We will consider a total of 17 different indicators, in five different domains: (1) incompleteness of ages/dates, four indicators; (2) inconsistent ages/dates, three indicators; (3) heaping or digit preference, four indicators; (4) transfer or displacement of dates, five indicators; and (5) over-dispersion of both the height-for-age and the weight-for-age *z*-scores, one indicator.

Only six of these indicators can be calculated for the five MIS surveys (Ghana, Kenya, Liberia, Nigeria, and Sierra Leone), because they did not include a survey of men and had only a truncated birth history. For those surveys, the only possible indicators are "IR_age_incomplete", "IR_age_shift", "hv105_heap_05", "PR_women_low_end", and "PR_women_high_end". The last two indicators, "BR_age_shift" and "U5_age_shift", cannot be calculated for the Mozambique AIS because it only included a truncated birth history and did not include measurements of height and weight.

Four indicators of incompleteness of age

Incompleteness of women's age: IR_age_incomplete. There are four points in DHS data where imputation or reconciliation of age is given. For all women, the data include month of birth, year of birth, and completed years of age at the time of the survey. If these are incomplete or inconsistent, then imputation rules are followed. Incompleteness is coded in detail, but in the great majority of cases the problem is just at the level of the month; the month of birth needs to be imputed or adjusted to achieve consistency. The indicator is coded 1 if v014>0; otherwise it is 0.

Incompleteness of women's age at first union: IR_marr_age_incomplete. Women are asked to provide a month, year, and age when first married or lived as married. The indicator is coded 1 if v510>0; otherwise it is 0.

Incompleteness of men's age: MR_age_incomplete. If included in the men's survey, men are asked to provide a month and year of birth and completed years of age at the time of the survey. The indicator is coded 1 if mv014>0; otherwise it is 0.

Incompleteness of children's age: BR_age_incomplete. Although the label refers to age, the incompleteness refers to the month and year of birth (b1 and b2), and it is assessed for each child in the birth histories collected for women, regardless of whether the child is still alive or is still living with the mother. The indicator is coded 1 if b10>0; otherwise it is 0.

Three indicators of inconsistent ages

Age change for women: IR_age_shift. The current age of women is assessed at two points. The first is in the household survey, simply in completed single years; the second is in the survey of women. These two responses are not automatically reconciled during data collection or computer editing—fortunately, because with automatic editing we would lose the opportunity for a consistency check. It is DHS policy to give priority to the age reported in the individual survey, because it is obtained directly from the woman herself and it involves detail on month and year of birth. The age in the household survey is reported by the household respondent, who is usually the household head or the spouse of the household head, and may be the woman herself, but discrepancies may be found even then. If the two ages differ by more than 2 years, the indicator is coded 1. Otherwise, it is 0.

Age change for men: MR_age_shift. If there is a survey of men, we can construct an analogous indicator of discrepancies between age in that survey and in the household roster. If the two ages differ by more than 2 years, the indicator is coded 1. Otherwise, it is 0.

Age change for children: BR_age_shift. The children in the birth histories, whether alive or dead, and whether living in the same household or not, are listed in the women's birth histories. The BR file consists of records for all such children, with the mother's data and all the health data attached. (The children born in the past 5 years are given in a shorter file, the KR file.) During data processing, a code is added giving the line number in the household file—if the child is alive and in the same household as the mother. Thus, for such children, it is possible to compare the years of age initially reported in the household survey with the years of age implied by the birthdate in the birth history. If the two ages differ by more than 2 years, the indicator is coded 1. Otherwise, it is 0.

Four indicators of age heaping

Age heaping of household members: hv105_heap_05. Age in the household file is given by hv105. In most surveys there is a tendency for ages to have 0 or 5 as the final digit. The indicator is coded 1 if the final digit is 0 or 5; otherwise it is coded 0. The indicator is restricted to ages 0-79; outside that range it is NA. The reason for this restriction is to have a range within which each of the 10 possible final digits appears the same number of times (in this case, eight times). In other data quality reports, age heaping has been assessed with Myers' Blended Index, which includes an adjustment for the overall downward gradient in the percentage of cases at ages 0, 1, …, 9. In this report, no blending is used. With accurate reporting of age, the digits 0 and 5 together should appear about 20% of the time.

Age heaping of women: v012_heap_05. The indicator is coded 1 if the final digit is 0 or 5; otherwise it is coded 0. The indicator is restricted to women age 20-49; outside that range it is NA. Within this range, each of the 10 possible final digits appears the same number of times (three times).

Age heaping of men: mv012_heap_05. The indicator is coded 1 if the final digit is 0 or 5; otherwise it is coded 0. The indicator is restricted to men age 20-49; outside that range it is NA. Within this range, each of the 10 possible final digits appears the same number of times (three times).

Age heaping of children in the birth histories: b8_heap_05. The indicator is coded 1 if the final digit is 0 or 5; otherwise it is coded 0. The indicator is restricted to children age 0-29; outside that range it is NA. Within this range, each of the 10 possible final digits appears the same number of times (three times).

Five indicators of age transfers

Women, low-end age transfers: PR_women_low_end. In many surveys there is some evidence of age displacement across a boundary for eligibility. For women, age 15-49 in the household survey is a criterion for eligibility for the women's survey. If a woman age 15 is misreported at age 14 in the household survey, she will not have to be included in the women's survey. Some surveys, however, show evidence of 14-year-olds being transferred to age 15 because of overtraining among interviewers to avoid the downward transfer. There are various ways to identify and measure such transfers, none of which is foolproof, because it is likely that some transfers cross more ages (e.g., from age 16 to age 13), and there are some genuine irregularities in the age distribution. We define the indicator to be 1 if a female in the household survey is reported at age 14, or 0 if she is reported at ages 13, 15, or 16. Outside the range 13-16, the indicator is NA. If there are no transfers, approximately one-fourth of the cases at age 13-16 will be reported at age 14. The measurement of low-end and high-end transfers for women and men is based on the PR file, i.e., the individual-level household data.

Women, high-end age transfers: PR_women_high_end. If a woman age 49 is misreported at age 50, she will not have to be included in the women's survey. We generally observe more evidence of transfers out of eligibility at the upper age, partly because older women will typically require a longer interview, and partly because age 50 tends to be over-reported just because of rounding or digit preference. We define the indicator to be 1 if a female in the household survey is reported at age 50, or 0 if she is reported at ages 48, 49, or 51. Outside the age range 48-51 the indicator is NA. If there are no transfers, approximately one-fourth of the cases at age 48-51 will be reported at age 50, but part of an excess above one-fourth is due to age heaping.

Men, low-end age transfers: PR_men_low_end. We define the indicator to be 1 if a male in the household survey is reported at age 14, or 0 if he is reported at ages 13, 15, or 16. Outside the age range 13-16 the indicator is NA. If there are no transfers, approximately one-fourth of the cases at age 13-16 will be reported at age 14.

Men, high-end age transfers: PR_men_high_end. The upper boundary of eligibility for men is not always age 49. We check for the possibility that it is a higher age, such as 54 or 59, and if it is, the following ages are adjusted appropriately. Apart from that modification, the indicator is 1 if a male in the household survey is reported at age 50, or 0 if he is reported at ages 48, 49, or 51. Outside the age range 48-51 the indicator is NA. If there are no transfers, approximately one-fourth of the cases at age 48-51 will be reported at age 50, but part of an excess above one-fourth is due to age heaping.

Birth year transfers: BR_1_2_3_4. In most DHS surveys, eligibility for the questions about child health begins with children born in January of the fifth calendar year before the beginning of fieldwork (the specific cmc is coded as v117 in the women's file; in some DHS surveys the starting month is a month other than January). Many surveys show evidence that some children's birthdates are pushed backward, making them ineligible for the health questions. Four years (not necessarily calendar years) are constructed and numbered 1, 2, 3, and 4. For example, if eligibility begins with births in January 2013, years 1 and 2 are the two years before the boundary, i.e., years 2011 and 2012, and years 3 and 4 are the two years past the boundary, i.e., years 2013 and 2014. We focus on transfers of births from year 3 to year 2, i.e., from 2013 to 2012. With appropriate modification depending on the survey, the indicator is 1 if a child's birth year is 2, or 0 if it is years 1, 3, or 4. Outside that 4-year period, the indicator is NA. If there are no transfers, approximately one-fourth of the births in that range will be coded 1.

One indicator of over-dispersion of age

Over-dispersion of children's ages: U5_age_disp. Most DHS surveys include the measurement of height and weight for all children under age 5 as part of the household survey. Many surveys also include such measurements for women age 15-49, and some include men age 15-49, but here we only consider children under age 5. The height and weight measurements are used to calculate three anthropometric z-scores, using WHO 2008 Child Growth Standards: HAZ (height for age), WAZ (weight for age), and WHZ (weight for height). Children are classified as stunted if HAZ<-2, underweight if WAZ<-2, and/or wasted if WHZ<-2. The HAZ scores are flagged with special codes if outside the range -6 to +6. The WAZ scores are flagged if outside the range -6 to +5, and the WHZ scores are flagged if outside the range -5 to +5. There is general agreement that over-dispersion in the z-scores is attributable to error in the measurement of one or more of the three variables that they are based on—height, weight, and age. Age errors are often overlooked as a potential source of over-dispersion in z-scores. Age errors could take the form of bias, i.e., a systematic tendency to report a child as too young, or as too old, at least within some part of the age range 0-4. They could also take the form of a tendency to add a random displacement up or down. We will use the proportion of children under age 5 who have flagged values for both the HAZ and WAZ[9] scores as an indirect measure of potential age misreporting. The indicator is 1 if both the HAZ and WAZ are flagged (that is, neither one is within the legal range), and 0 otherwise, except that the indicator is NA if the measurement is NA (for example, if the survey did not include anthropometry). The measurements of height and weight are taken by the biomarker specialists, but errors in the measurement of age would be traced to the woman's interviewer, who is responsible for the information in the birth history. This indicator represents an indirect strategy to identify interviewers' uncertainty and inaccuracy when estimating the age of young children.

6.2 Identifying Problematic Surveys and Indicators

We have examined the prevalence and variation of all indicators in all surveys, but rather than giving complete detail, only a few generalizations and specific examples will be provided within this report.

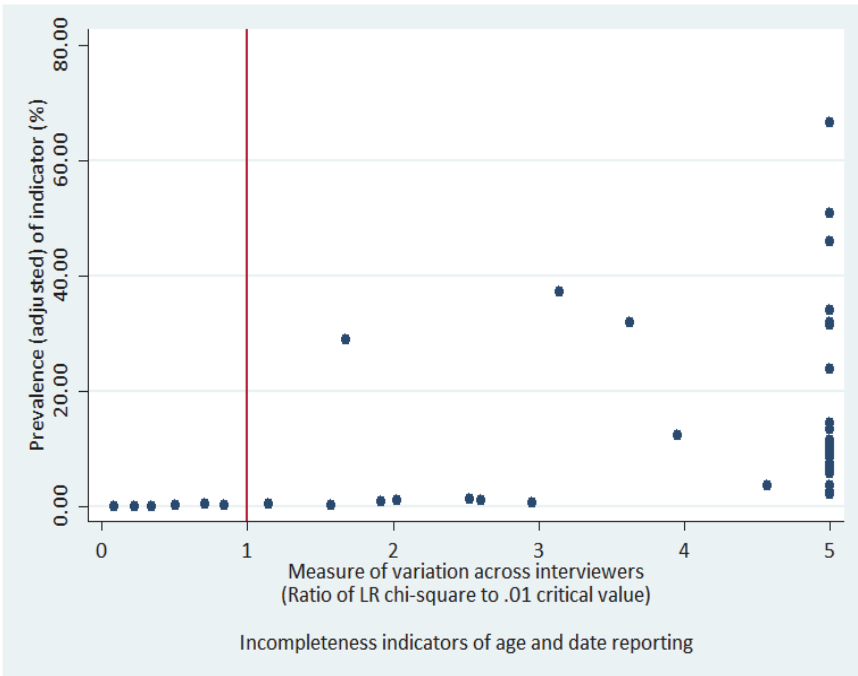
In this section we review the five domains of age and date indicators in terms of identifying problematic combinations of surveys and indicators—problematic in the sense of having relatively high prevalence of a symptom and also having highly significant variation across interviewers. For each domain, we provide a scatterplot that includes all the indicators in that domain, for an overview. Numerical values are given in tables in Appendix 3. In all five scatterplots, a dot refers to a combination of an indicator within the domain and a survey. The horizontal axis is the chi2 ratio defined earlier—that is, the ratio of the chi-square statistic D to its critical value for a .01 test of the null hypothesis that there is no variation across interviewers. If a dot is to the right of the vertical red line, then the ratio is greater than 1 and the null hypothesis would be

[9] Prior to DHS-7, all three Z scores would be flagged for the same cases. In recent surveys, there is overlap but each Z score is flagged separately.

rejected at the .01 level. If the ratio is greater than 5, it is replaced with 5. The vertical axis is the prevalence of the outcome, that is, the overall percentage of responses (for the specific indicator in a specific survey) that have code "1". The vertical axis varies considerably from one domain to another.

Incompleteness of age

Figure 6.1 Combinations of country and indicators of incompleteness of age for which variation across interviewers is significant at the .01 level. The horizontal scale is the ratio of the calculated chi-square to the critical value for a .01 test of significance.



Among the indicators of incompleteness of age, the one most often problematic is women's age at marriage. The level ranges from a low of 23.81% in Nepal to a high of 66.43% in Mozambique. There is highly significant variation across interviewers for this indicator in every survey.

Incompleteness of age in the birth histories is not problematic. The highest prevalence is in Malawi, at only 2.67%, followed by 2.24% in Uganda. In both cases, there is highly significant variation across interviewers, with Chi2 ratios of 6.73 and 6.25, respectively, but the prevalence is low enough that we would not suggest further analysis.

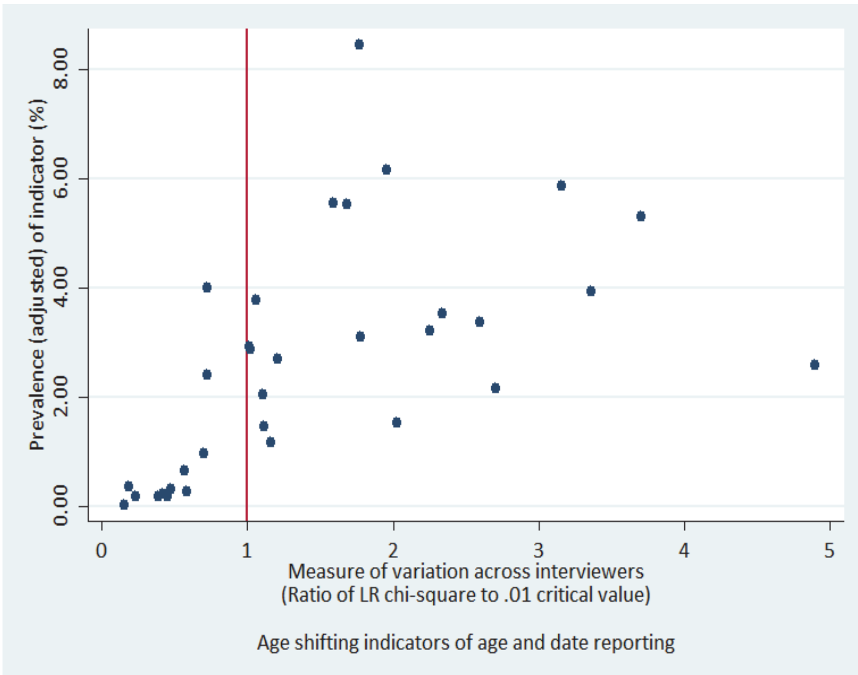
The levels of incompleteness of age for women and men are similar. The overall level is 6.87% for women's age and 6.08% for men's age. The prevalence is similar for women and men within surveys, as well. Women's and men's interviews are almost always conducted by different interviewers, but women and men on the same team have the same supervisor, and within a specific country women and men are trained together.

However, these similarities mask a great deal of variation across surveys and, within surveys, across interviewers. In the Armenia, Timor-Leste, and Zimbabwe surveys, for both women's age and men's age the prevalence is less than 2%, and in Armenia and Zimbabwe the deviance across interviewers is not significant. Otherwise, the prevalence is 3% or more and the deviance across interviewers is highly significant. Section 6.3 will examine inconsistent reporting of women's age in the Kenya survey, which at 14.52% has the highest prevalence of incompleteness for women or men.

In terms of potential diagnostic value, the most useful indicators of age incompleteness are those for women's age and men's age; the latter is only relevant when there is a survey of men.

Inconsistent ages

Figure 6.2 Combinations of country and indicators of inconsistent ages for which variation across interviewers is significant at the .01 level. The horizontal scale is the ratio of the calculated chi-square to the critical value for a .01 test of significance.



If a woman's age in the household survey differs from her age in the women's survey by 2 or more years, and similarly for men and for children, then we say that the age has been shifted or the two ages are inconsistent.

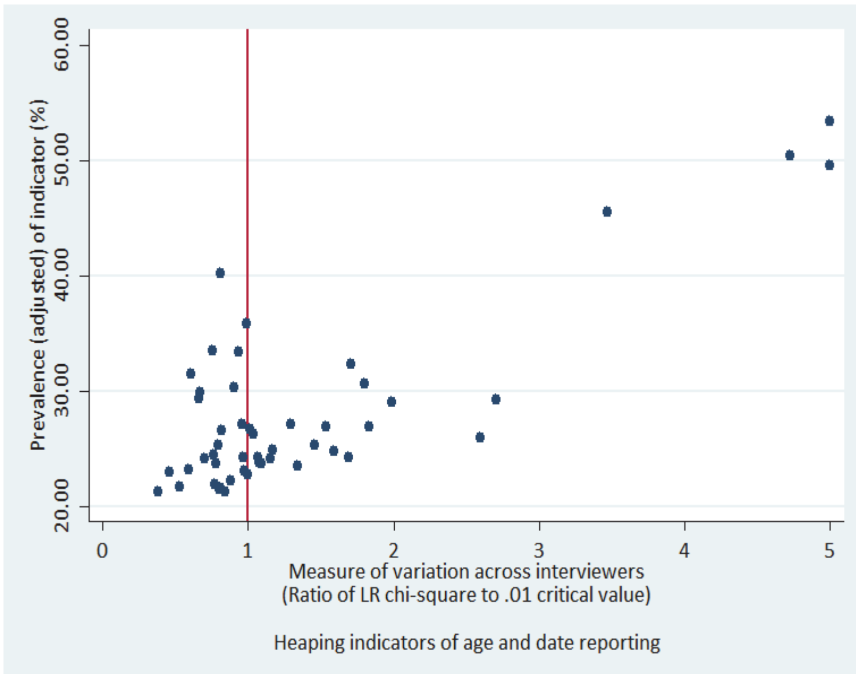
For children in the birth histories who are also in the household survey, the ages in the two sources agree closely. The prevalence of inconsistencies is always well below a threshold level of 2%. Consistency is only an issue for women and men and, again, the levels for women and men are remarkably similar. The overall level for both is 3.4% (3.35% for women and 3.37% for men).

In a majority of surveys, the level of inconsistency for women or men (comparing age in the individual interview with age in the household survey) is above a 2% threshold. The highest level is 8.46% for women in the Uganda survey. That survey will be examined more closely in Section 6.3.

In terms of potential diagnostic value, the most useful indicators of age inconsistencies, as for age incompleteness mentioned above, are those for women's age and men's age; the latter is only relevant when there is a survey of men.

Age heaping

Figure 6.3 Combinations of country and indicators of age heaping for which variation across interviewers is significant at the .01 level. The horizontal scale is the ratio of the calculated chi-square to the critical value for a .01 test of significance.



We would expect all of the age heaping indicators to have a prevalence of about 20%, simply because 0 and 5 represent 20% of the digits 0 through 9. The mean that is observed (across all four indicators and all surveys, without adjusting for the number of cases) is 28.21%, well above 20%. The indicator is highest for children in the birth histories (32.55%), second highest for household members in the household survey (30.27%), third highest for women in the survey of women (27.00%), and lowest for men in the survey of men (23.20%).

When the age in the household survey is revised in the individual survey (a change that was described above as an inconsistency), the revision may tend to be an improvement in terms of closer correspondence with the respondent's true age. However, this is not necessarily implied by the finding that there is less heaping in the surveys of women and men than in the household survey. Much of the heaping in the household survey is due to ages above the upper end of age range for eligibility in the surveys of women and men.

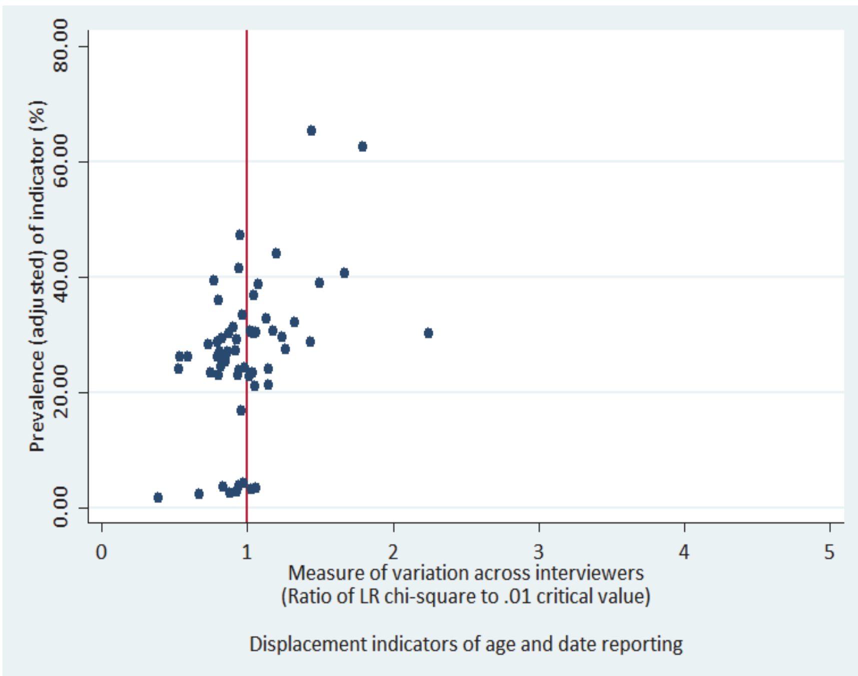
The overall prevalence is higher for women than for men, but within specific surveys there is little difference. Nigeria and Survey Leone had extremely high levels of heaping in their surveys of women—50.34% in Nigeria and 45.51% in Sierra Leone. Because those two surveys did not include a survey of men, the overall average for men was reduced to a lower level than that for women. In Nigeria and Sierra Leone, heaping in the household survey was even more severe than in the survey of women.

The highest prevalence (53.40%) and the highest Chi2 ratio (9.75) were for age heaping in the household survey in Nigeria. The problem was not shared among a handful of interviewers, but was endemic. Details for this combination are explored in Section 6.3.

The most useful diagnostic indicator of age misreporting in this group, on heaping or digit preference, is heaping on final digits 0 and 5 in the household survey.

Age transfers

Figure 6.4 Combinations of country and indicators of age displacement for which variation across interviewers is significant at the .01 level. The horizontal scale is the ratio of the calculated chi-square to the critical value for a .01 test of significance.



The five indicators of displacement would be expected to have a prevalence of approximately 25%, because they are constructed to identify a preponderance of cases in one year of age (or, in the case of birth history transfers, one year of time) out of four. Interpretation of the observed prevalence must take that into account. The interpretation of the deviance does not depend on that standard, because variation is always relative to the observed mean.

Across all surveys, the prevalence is 26.98% for the low end of the age range for women, 36.99% for the high end for women, 26.81% for the low end for men, 31.28% for the high end for men, and 28.09% for the birth history transfers. All of these levels are greater than 25%; overall, there is a tendency for all types of transfers to occur. However, the overall prevalence of the downward transfers is within a 2% difference from the expectation, and the birth transfers are only slightly more prevalent.

Upward transfers appear to occur about 5 percentage points more often than downward transfers for men, and about 10 percentage points more often than downward transfers for women. The relatively high prevalence of the upward transfers for women is largely due to the fact that age 50, a multiple of 10, is the destination age for those transfers, and there is generally a good deal of heaping at 50 that cannot be distinguished from transfers. The upper age of eligibility for men is not always 49.

The overall prevalence of downward age transfers is almost exactly the same for women and men. Within countries, it is also similar.

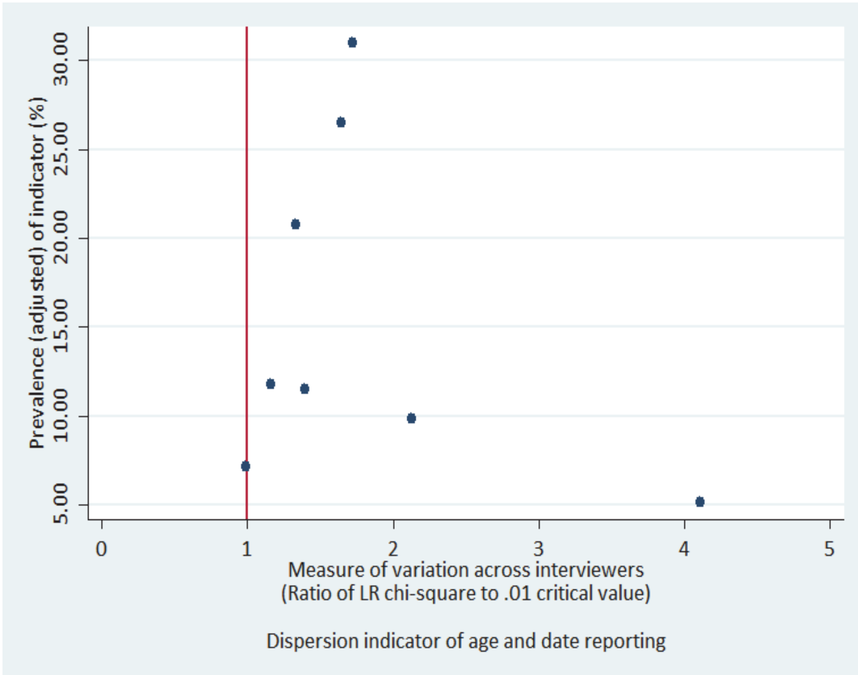
There are combinations of survey and indicator that are problematic in the sense of the prevalence being substantially above 25%, or being below 25%, a deviation that suggests overcorrection of the tendency to transfer ages outside the range of eligibility. However, the focus of this report is on variation across interviewers. The deviance D for interviewer effects is almost always significant at the .01 level for women's upward age transfers. That is, Chi2 ratio is almost always greater than 1; however, it is never more than 1.8. The other indicators of age transfers do not usually show significant variation across interviewers.

A combination that stands out, with one of the largest deviations from 25%, and also the largest D by far, is downward age transfers for women in Sierra Leone. The prevalence is 30.26%, and the Chi2 ratio is 2.24. This combination will be explored in Section 6.3.

The most useful indicator in this group for diagnostic purposes appears to be downward age transfers for women. Potential age transfers in the birth histories are important and should be monitored, but they appear to have less value for identifying interviewer effects. The deviances across interviewers are marginally significant in Tanzania and Timor-Leste (with Chi2 ratios or 1.06 and 1.03 respectively), but not in the other surveys.

Over-dispersion of children's age

Figure 6.5 Combinations of country and the indicator of over-dispersion of children's age for which variation across interviewers is significant at the .01 level. The horizontal scale is the ratio of the calculated chi-square to the critical value for a .01 test of significance.



The indicator of over-dispersion of age for children under age 5, which has not been used in previous data quality analyses, is based on over-dispersion of the z-scores for height-for-age (WAZ) and weight-for-age (HAZ). If both of those z-scores are flagged, then the reason may be because age has been poorly measured.

The indicator can only be calculated for surveys that include anthropometric measurements. In this report, the relevant surveys are Armenia, Burundi, Myanmar, Malawi, Nepal, Timor-Leste, Tanzania, Uganda, and Zimbabwe.

The indicator ranges from a low of 5.14% in Armenia to a high of 31.01% in Timor-Leste. Chi2 ratio is greater than 1 in all surveys except Zimbabwe, where it is 0.99, just barely below 1. Section 6.3 includes more detail on Timor-Leste, where Chi2 ratio for the interviewers is 1.72.

Because this indicator generally produces statistically significant evidence of interviewer effects, we suggest that it indeed has diagnostic value. The interpretation of the indicator in terms of age needs development. It may indirectly refer to the measurement of children's ages, but it is drawn from the anthropometric scores and refers more directly to the measurement of weight and height.

6.3 Measuring the Concentration of Problematic Interviewers

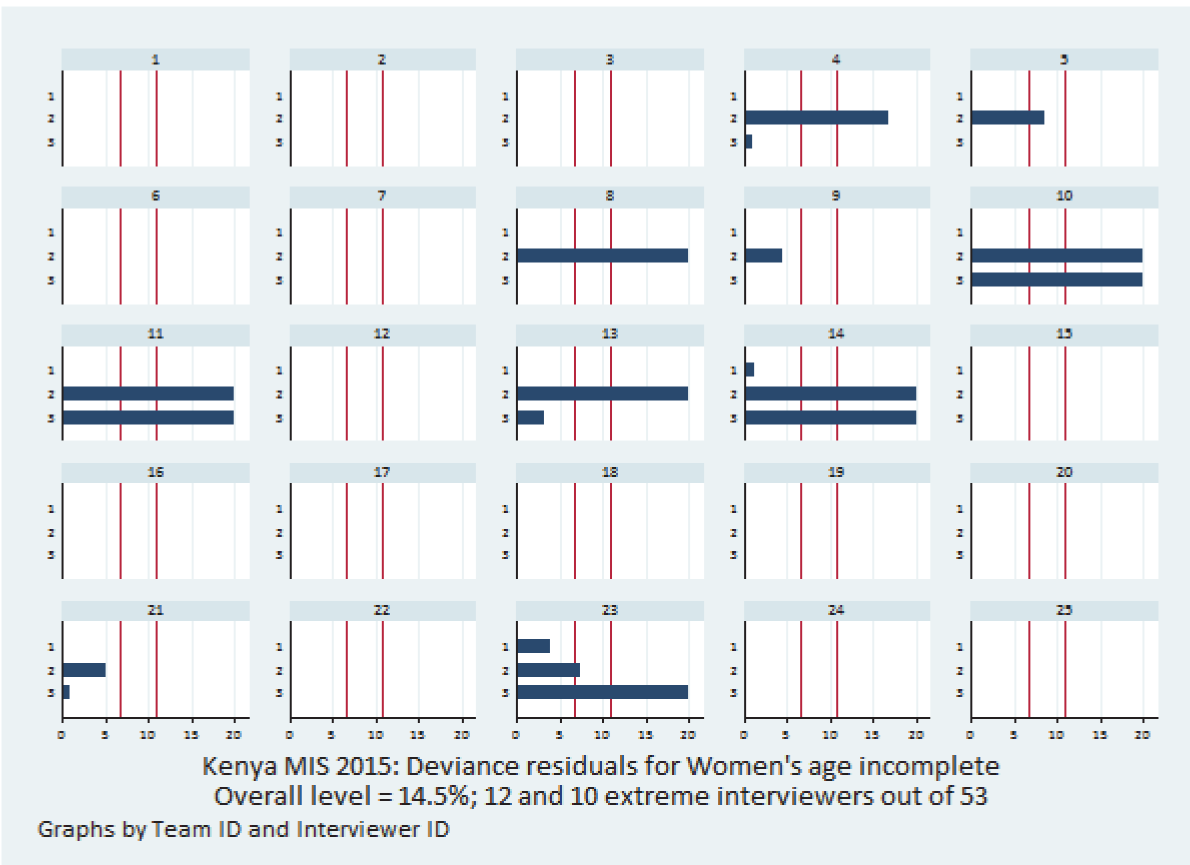
Measures of concentration and figures such as those in Sections 4.3 and 5.3 have been prepared but are not included. For each of the five types of misreporting of age, a specific example is explored in detail.

Incompleteness of age

Incomplete reporting of women's age in the Kenya survey is an example of a relatively high prevalence, at 14.52%, combined with a very high deviance D across interviewers. The ratio of the deviance to its .01 critical value is 16.8 (the deviance is 1,323.12, with 52 degrees of freedom. The horizontal bar graph identifying interviewers with large positive deviance residuals is shown in Figure 6.1. Out of 53 interviewers, 10 have squared deviance residuals dr2, for higher than average incompleteness, that are significant at the .001 level. Figure 6.1 shows that 6 of the 10 are on three teams, teams #10 #11, and #14.

List 6.1 provides more detail, listing the complete diagnostics for the interviewers who are most problematic, as well as others whose deviance residuals are significant at the .05 level. We note that in this survey, 14 interviewers conducted a total of 1,494 interviews with not a single incomplete age for women. For the 10 most problematic interviewers, 30% or more of their reported ages of women were incomplete. When those 10 interviewers are dropped, the mean prevalence drops by 44%, to 8.15%.

Figure 6.6 Deviance residuals dr2 for interviewers who are above the mean, for incompleteness of women's age in Kenya, by interviewer within team



List 6.1 Interviewer-level distribution of incompleteness of women's age in the Kenya survey, with deviance residuals. Interviewers with dr<=1.96 are omitted.

....

id	n1	n0	n	p1	dr	dr2
92	26	93	119	21.85	2.14	4.58
212	25	86	111	22.52	2.25	5.05
232	26	80	106	24.53	2.71	7.35
52	28	84	112	25.00	2.91	8.46
42	33	78	111	29.73	4.09	16.74
82	31	61	92	33.70	4.61	21.21
143	26	12	38	68.42	7.53	56.72
113	38	34	72	52.78	7.60	57.75
132	37	28	65	56.92	7.92	62.74
233	42	36	78	53.85	8.11	65.74
102	55	57	112	49.11	8.66	74.94
112	42	23	65	64.62	9.21	84.85
103	58	41	99	58.59	10.12	102.42
142	46	14	60	76.67	10.81	116.75

Inconsistent ages

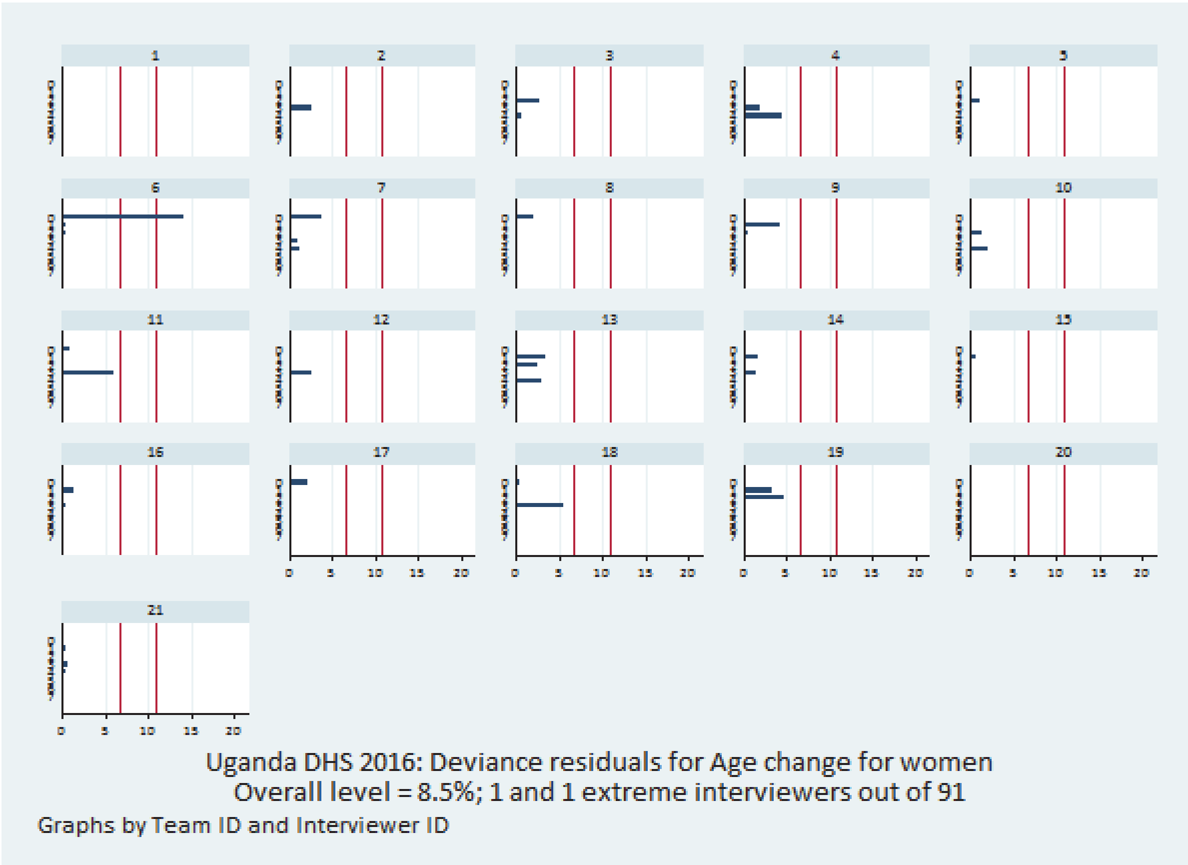
Interviewer effects on inconsistent ages will be illustrated with the survey in Uganda, which has the highest level for any of the three indicators, at 8.46% for the inconsistency of women's age in the individual and household surveys. The bar graph and list of interviewer numbers are given in Figure 6.2 and List 6.2, respectively. The list is truncated to show only the interviewers with dr>1.96.

The deviance D across interviewers was 219.11, with 90 degrees of freedom, a value that is 1.77 times the .01 critical value for a chi-square statistic with 90 degrees of freedom. As Figure 6.2 shows, out of the 91 interviewers there was only one, #600,[10] with a highly significant positive deviation. This interviewer was a clear outlier in the distribution of the prevalence of this type of inconsistency. List 6.2 shows that interviewer #600 did 140 interviews of households and women and in 26 cases the age in the household interview and the age in the women's interview differed by more than 2 years. That is, her prevalence for this outcome was 18.57% compared with an overall prevalence of 8.46%—more than twice the overall prevalence. Apart from interviewer #1004, who did only 15 interviews, of which 3 (20%) had an inconsistency, the next highest prevalence was for interviewer #1902, who had a prevalence of 13.37% in 172 interviews. Interviewer #600 was clearly an outlier. Of a total positive deviance (that is, the part of the deviance D that was due to a higher than average prevalence of inconsistencies) of 86.38, 14.19 was due this this single interviewer. That is, about 16% of the problematic deviance was due to one interviewer out of 91.

Despite the clearly excessive level of inconsistencies for this interviewer, dropping her alone does not produce an "acceptable" deviance D. If she is dropped, the overall prevalence drops only slightly, to 8.39%, and the deviance declines only slightly, to 204.78, which is still highly significant; Chi2 ratio declines to 1.67. There are many interviewers with prevalence of 10% to 12%, scattered across different teams, as well as many interviewers with prevalence well below the mean of 8.46%. For this indicator, the highly significant variation across interviewers cannot be reduced to one or just a small number of problematic individuals or problematic teams.

[10] The ID codes for interviewers in the Uganda survey were constructed with a "0" between the team number and the single digit for each member of the team. Thus, the ID code for person 0 on team 6 is 600.

Figure 6.7 Deviance residuals dr2 for interviewers who are above the mean, for inconsistency of women's age in Uganda, by interviewer within team



List 6.2 Interviewer-level distribution of inconsistency of women's age in the Uganda survey, with deviance residuals. Interviewers with dr<=1.96 are omitted.

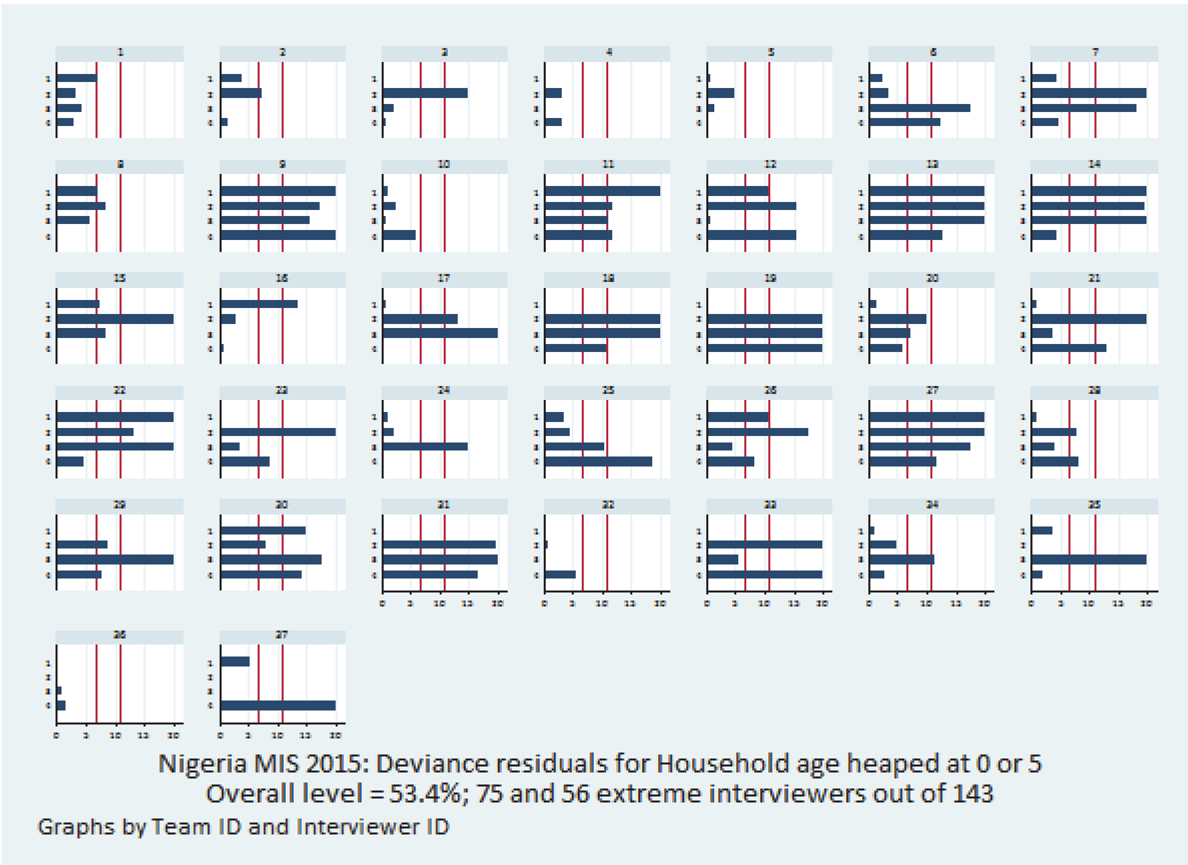
Id	n1	n0	n	p1	dr	dr2
901	29	204	233	12.45	2.05	4.22
404	31	219	250	12.40	2.10	4.43
1902	23	149	172	13.37	2.15	4.62
1803	32	217	249	12.85	2.33	5.42
1103	36	245	281	12.81	2.45	6.01
600	26	114	140	18.57	3.77	14.19

Age heaping

The preceding section showed that the most severe instance of heaping was in the Nigeria household survey, with a prevalence of 53.40%, and this was also the combination with the highest deviance across interviewers, D=9.75. We now examine the distributions across interviewers for this indicator and survey, which was a large survey with 143 interviewers. The bar graph for interviewers is given as Figure 6.3, and the frequencies are in List 6.3. In the graph, the bars for dr2 are truncated at 20.

There is an extraordinary amount of variation in the level of heaping in this survey. Only two interviewers had a prevalence below 20%; there was clearly not the kind of overcorrection or avoidance of 0 and 5 that was observed in some other surveys. The interquartile range for the prevalence, the range within which the middle half of all interviewers were located, goes from 38.40% to 63.73%—that is, from approximately two times to approximately three times what the correct prevalence should be. There is no clear boundary between interviewers whose level of heaping is moderate, say, and those whose level is high. Dozens of interviewers are simply heaping far more often than would be acceptable by any standard.

Figure 6.8 Deviance residuals dr2 for interviewers who are above the mean, for heaping on final digits 0 and 5 in the Nigeria household survey, by interviewer within team



List 6.3 Interviewer-level distribution of heaping of age at final digits 0 or 5 in the Nigeria household survey, with deviance residuals. Interviewers with dr<=1.96 are omitted.

id	n1	n0	n	p1	dr	dr2
1404	65	37	102	63.73	2.11	4.44
704	79	46	125	63.20	2.21	4.90
502	89	53	142	62.68	2.23	4.98
2004	87	50	137	63.50	2.39	5.71
803	70	38	108	64.81	2.40	5.76
1004	32	13	45	71.11	2.43	5.90
2003	117	68	185	63.24	2.70	7.31
1501	6	0	6	100.00	2.74	7.53
2002	83	40	123	67.48	3.17	10.06
1201	70	31	101	69.31	3.26	10.61
1804	98	49	147	66.67	3.26	10.65
1103	121	64	185	65.41	3.31	10.94
1102	108	54	162	66.67	3.43	11.74
1104	85	39	124	68.55	3.43	11.78
604	78	34	112	69.64	3.50	12.28
1304	79	34	113	69.91	3.58	12.81
1702	111	54	165	67.27	3.62	13.10
903	82	33	115	71.30	3.93	15.41
1202	97	42	139	69.78	3.94	15.51
1204	89	37	126	70.63	3.95	15.61
902	97	40	137	70.80	4.16	17.31
603	89	35	124	71.77	4.19	17.53
703	102	42	144	70.83	4.27	18.26
1402	109	45	154	70.78	4.41	19.41
1301	117	49	166	70.48	4.49	20.19
901	97	37	134	72.39	4.50	20.28
1703	126	53	179	70.39	4.64	21.53
1403	101	37	138	73.19	4.77	22.76
1101	76	22	98	77.55	4.96	24.57
1302	114	41	155	73.55	5.15	26.54
1904	76	17	93	81.72	5.73	32.85
702	112	31	143	78.32	6.19	38.33
904	116	32	148	78.38	6.31	39.87
1303	89	18	107	83.18	6.50	42.19
1903	167	51	218	76.61	7.09	50.22
1502	136	35	171	79.53	7.12	50.74
1401	126	22	148	85.14	8.20	67.25
1803	202	53	255	79.22	8.59	73.71
1902	160	26	186	86.02	9.48	89.94
1802	156	22	178	87.64	9.81	96.16

Age transfers

In the Sierra Leone household survey, 30.3% of women's ages reported at 13, 14, 15, or 16 were reported at age 14. This is substantially greater than a level of about 25% that would have been expected in the absence of misreporting. This net downward shift or displacement out of the age range for eligibility for the individual interview masks a great deal of variation across the 80 interviewers.

Earlier surveys in Sierra Leone showed strong evidence of downward age transfers, and attempts were made to avoid this type of displacement in the 2016 survey. The interviewers were trained specifically to avoid under-reporting of age 15 and over-reporting of age 14. Unfortunately, the result was displacement in the other direction, shifting girls upward into the age range of eligibility. This is seen clearly by looking at the numbers of interviewers whose prevalence was below or above a range from 20% to 30%, that is, 25% +/- 5%. Twenty-three interviewers appear to have overcorrected for potential age transfers, with a prevalence below 20%, and 36 had a prevalence above 30%, while 21 were within the range of 20% to 30%.

The overall deviance of 249.48 (with 79 degrees of freedom) is heavily affected by the interviewers who overcorrected, as well as by those who shifted ages upward, in the direction that reduced their workload. The 23 interviewers with a prevalence below 20% account for 138.07 points of deviance. The 36 with prevalence above 30% account for 100.30 points of deviance. The 21 in the range from 20% to 30% account for only 11.11 points of deviance. In other words, in terms of dispersion across interviewers, downward transfers were more of a problem than upward transfers, even though the upward transfers dominated in terms of the combined or net prevalence.

Figure 6.4 shows the deviance residuals dr2 that includes both the downward and upward deviations. For example, interviewer #802[11] had a large negative deviation that accounted for 25.19 points of deviance (truncated to 20 in the figure). Interviewer #1203 had a large positive deviation that accounted for 15.75 points of deviance. Eleven interviewers were significant at the .01 level and six of them were significant at the .001 level. Of those six interviewers, four were significantly low (with downward transfers) and two were significantly high (with upward transfers).

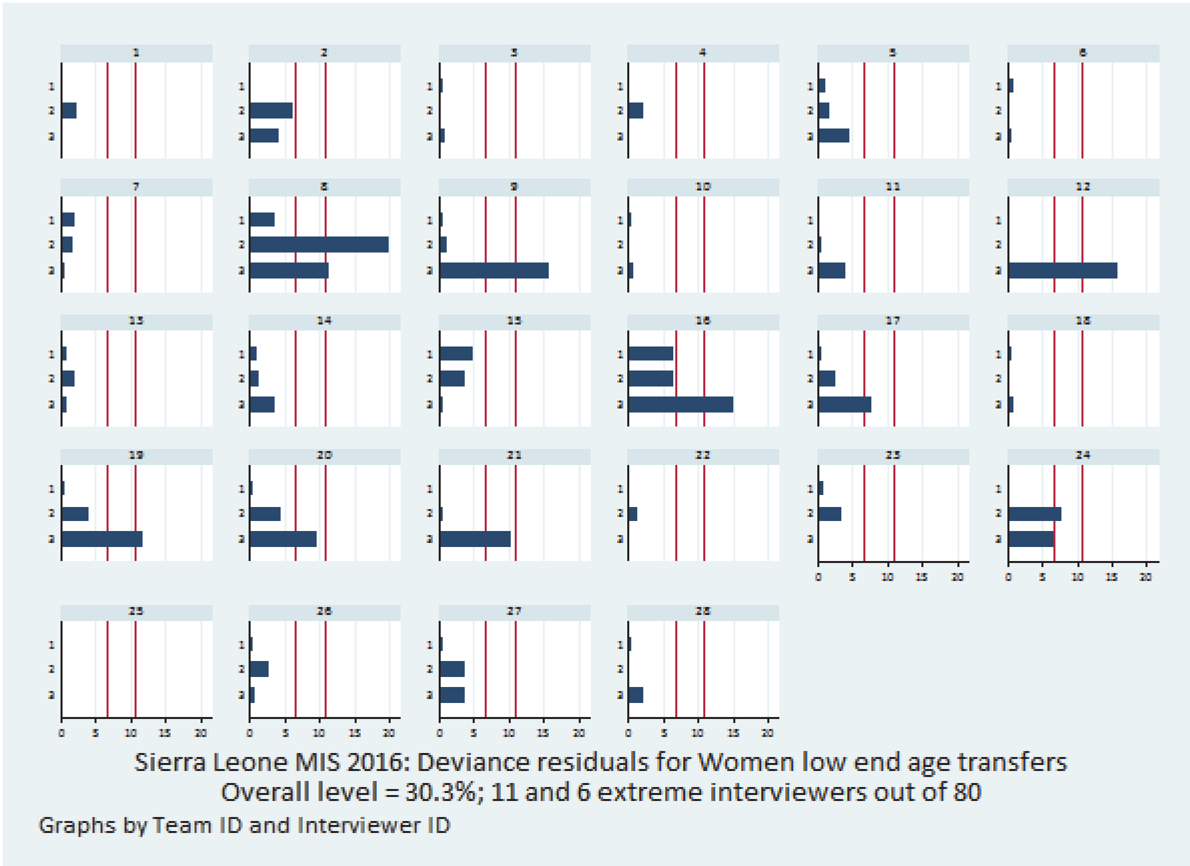
The direction of significant deviations from the overall prevalence was generally the same within a team, regardless of whether the deviations were statistically significant. There is only one team, #24, within which one team member (#2402) was significantly low and another team member (#2403) was significantly high. It appears that the avoidance of transfers from age 15 to 14, with ensuing transfers in the opposite direction, was enforced with supervision in several teams, particularly #2, #8, and #16, and by contrast was virtually ignored in other teams, particularly #15, #19, and #20.

List 6.4 provides detail on this combination of indicator and survey. The middle range of interviewers, whose value of dr was in the range of +/-1.96, is omitted from the list. If the six most extreme interviewers—the four at the low end and the two at the high end for whom dr2 is greater than the .001 critical value—are dropped, then Chi2 ratio declines by a third, from 2.24 to 1.48, and the prevalence of the indicator rises slightly, from 30.26% to 31.10%. Nevertheless, the variation across interviewers remains highly significant.

The interviewer effects for age transfers in the Sierra Leone survey, like the interviewer effects for heaping in the Nigeria survey discussed above, are only partially attributable to a relatively small number of interviewers who were inadequately trained or supervised. Rather, the variation across interviewers appears to be due to a lack of training and supervision for several teams, which were free to make upward transfers in large numbers, combined with selective training and supervision for several other terms that inhibited or penalized the use of age 14, specifically. This inconsistency across teams and the evident pressure to avoid age 14, described here as "overtraining," are sources of variation that are distinct from the measured characteristics of the interviewers.

[11] The ID codes for interviewers in the Sierra Leone survey were constructed with a "0" between the team number and the single digit for each member of the team. Thus, the ID code for person 2 on team 8 is 802.

Figure 6.9 Deviance residuals dr2 for interviewers, for downward displacement of women's age in the Sierra Leone household survey, by interviewer within team



List 6.4 Interviewer-level distribution of the indicator of downward displacement of women's age in the Sierra Leone survey, with deviance residuals. Interviewers with dr>=-1.96 or dr<=1.96 are omitted.

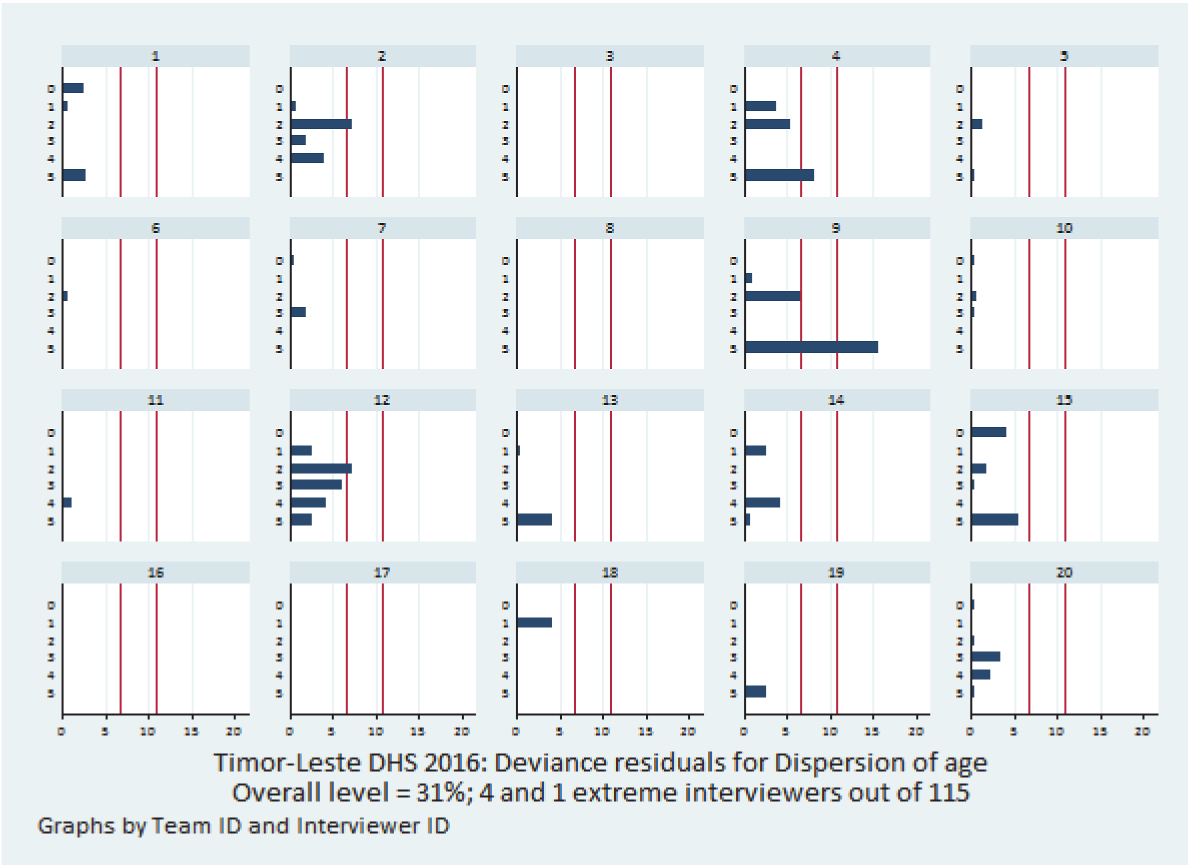
id	n1	n0	n	p1	dr	dr2
802	1	45	46	2.17	-5.02	25.19
903	0	22	22	0.00	-3.98	15.86
1603	2	36	38	5.26	-3.88	15.06
803	1	24	25	4.00	-3.36	11.29
2103	2	28	30	6.67	-3.20	10.27
2402	0	11	11	0.00	-2.82	7.93
1601	0	9	9	0.00	-2.55	6.49
202	2	21	23	8.70	-2.52	6.33
1602	2	21	23	8.70	-2.52	6.33
503	1	13	14	7.14	-2.13	4.56
203	2	17	19	10.53	-2.06	4.25
....						
1103	15	17	32	46.88	1.97	3.88
1502	15	17	32	46.88	1.97	3.88
1902	20	25	45	44.44	2.00	4.01
2002	15	16	31	48.39	2.11	4.45
1501	2	0	2	100.00	2.19	4.78
2403	13	10	23	56.52	2.61	6.80
1703	21	20	41	51.22	2.79	7.81
2003	19	15	34	55.88	3.09	9.57
1903	15	8	23	65.22	3.45	11.91
1203	30	23	53	56.60	3.97	15.75

Over-dispersion of children's age

The proposed indicator of measurement error in children's age would take the value 0 in the absence of flagging of the HAZ and WAZ anthropometric scores. In the Timor-Leste survey, the overall value is a very high 31.01%, with a total deviance D across interviewers of 262.25 and a Chi2 ratio of 1.72. Figure 6.5 and List 6.5 are limited to the interviewers with a positive deviance, but even those with a negative deviance—that is, those with a prevalence below the overall prevalence—are well above the theoretical minimum of 0. No interviewer has a prevalence below 14%.

The variation in prevalence across interviewers is indeed substantial, as high as 51%. Interviewer #95, who had the highest prevalence, contributed dr2=15.66 points to the total deviance. List 6.5 includes 13 other interviewers whose prevalence was in the range of 41% to 49%. Interviewer #95 was only 2 percentage points higher. Because of the large number of interviewers in this survey, at 115, a simulation that removes this single interviewer does not produce an appreciable reduction in the total deviance.

Figure 6.10 Deviance residuals dr2 for interviewers for over-dispersion of children's ages, inferred from over-dispersion of the HAZ and WAZ scores in the Timor-Leste survey, by interviewer within team



List 6.5 Interviewer-level distribution of the indicator of over-dispersion of children's ages in the Timor-Leste survey, with deviance residuals. Interviewers with dr<=1.96 are omitted.

id	n1	n0	n	p1	dr	dr2
150	26	34	60	43.33	2.00	4.02
24	20	24	44	45.45	2.00	4.02
135	29	39	68	42.65	2.02	4.07
181	22	27	49	44.90	2.04	4.15
144	40	58	98	40.82	2.05	4.20
124	24	30	54	44.44	2.07	4.28
42	45	63	108	41.67	2.33	5.44
155	25	29	54	46.30	2.35	5.51
123	25	28	53	47.17	2.45	6.03
92	35	43	78	44.87	2.56	6.57
22	25	26	51	49.02	2.68	7.16
122	31	35	66	46.97	2.71	7.32
45	44	54	98	44.90	2.88	8.29
95	46	44	90	51.11	3.96	15.66

6.4 Associations with Interviewer Characteristics

This section will examine the relationship between interviewer characteristics and the indicators of errors in the measurement of ages and dates. It would be desirable to reduce the 17 potential indicators defined in Section 6.1 to a single composite indicator. However, the patterns of variations are not similar enough to justify pooling. Instead, one diagnostic indicator will be selected from each of the five groups, as follows:

Incompleteness of women's age, in the individual survey
Inconsistent ages of women, in the household and individual surveys
Heaping of ages ending in 0 or 5, in the household survey
Displacement of women's ages across the lower age of eligibility, in the household survey
Over-dispersion of the ages of children under age 5, in the household survey

For each of these five indicators we provide two figures showing 95% confidence intervals for the covariates, and a table. The figures readily identify the different patterns. The tables give numerical values and also indicate the significance of the variation between categories. Unless otherwise indicated, all covariates are significant at the .01 level, and generally at a much higher level. The significance of a covariate refers to whether the means for the categories are different from one another. The figures and tables include a "missing" category and the overall level, for easier visual reference. The "missing" category is often significantly different from the other categories, but tends to have a wide confidence interval because it contains fewer cases than the other categories. All the subgraphs include a vertical line at the overall mean for the indicator, making it easier to identify categories that are above or below the mean.

The first of the two figures includes the interviewer's place of residence, age group, sex, and marital status. The second includes the interviewer's education, previous experience with DHS surveys, previous experience with other surveys, and whether she/he is a permanent employee of the implementing agency.

The units of analysis are interviews, not interviewers—although it would be equivalent to say that the units are interviewers weighted by the number of interviews they conducted. The representation of each survey is proportional to the number of relevant cases in that survey. The number of cases is given in the bottom row of each table. For all the outcomes, a higher percentage or prevalence is interpreted as being more problematic. Thus the categories with the lowest prevalence, shown farthest to the left in the figures, have the highest data quality.

Incompleteness of women's age

Figure 6.11a Point estimates and the lower and upper ends of 95% confidence intervals for the prevalence of incompleteness of women's age, within categories of the interviewer's place of residence, age group, sex, and marital status. All surveys are pooled. See Appendix Table A3.3 for numerical values.

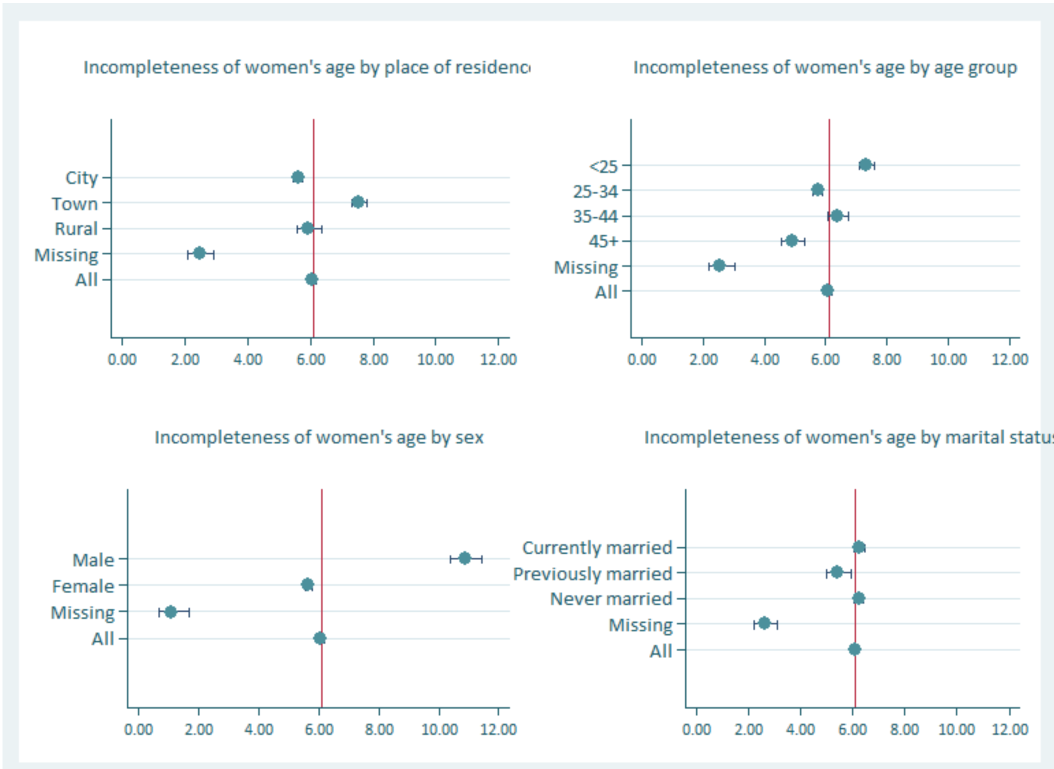
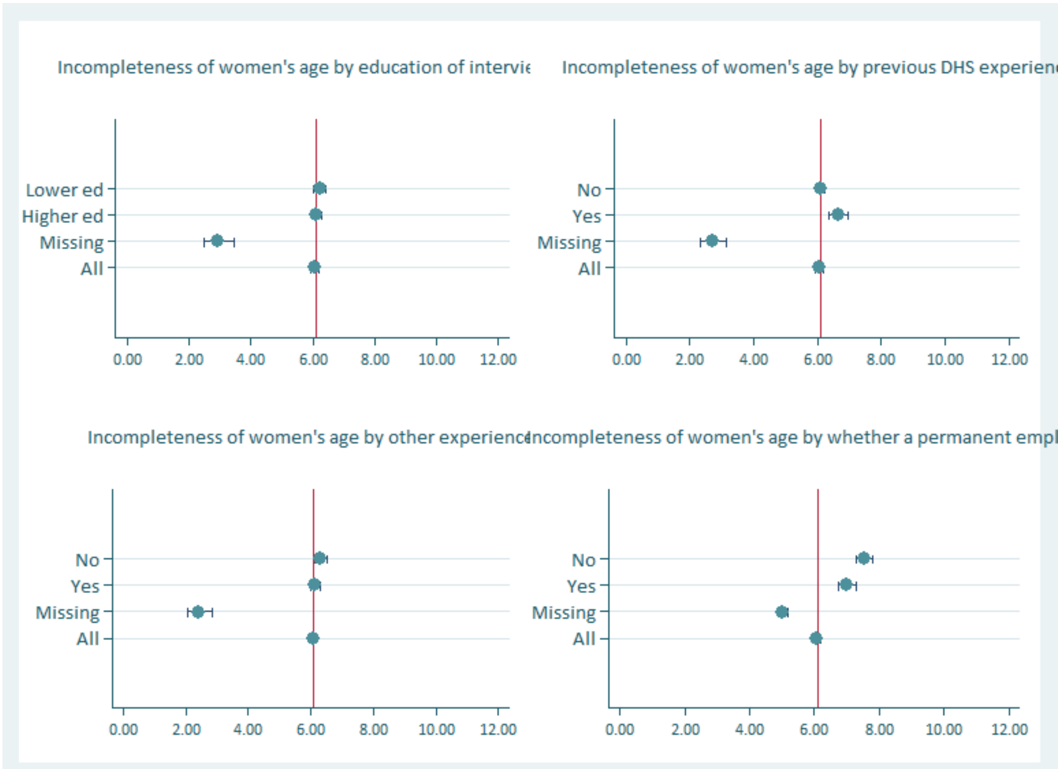


Figure 6.11b **Point estimates and the lower and upper ends of 95% confidence intervals for the prevalence of incompleteness of women's age, within categories of the interviewer's education, previous DHS experience, previous experience with other surveys, and whether a permanent employee of the implementing agency.** All surveys are pooled. See Appendix Table A3.3 for numerical values.



Place of residence: The only significant difference is that "town" interviewers have higher levels of incompleteness than "city" or "rural" interviewers.

Age group: Incompleteness is highest for the youngest interviewers and lowest for the oldest interviewers. The two intermediate age groups are in the middle.

Sex: Female interviewers have much lower levels of incompleteness than male interviewers. The difference is approximately 4 percentage points.

Marital status: Previously married interviewers have significantly lower incompleteness than the other two categories, but the difference is only about 1 percentage point.

Education: There is no significant difference between the two education categories. The "missing" cases have the lowest levels of incompleteness.

Previous DHS experience: Interviewers with previous DHS experience have somewhat higher levels of incompleteness compared with interviewers with no DHS experience. The difference is small, about half a percentage point, but significant.

Other survey experience: There is no significant difference between the two categories of experience. The "missing" cases have the lowest levels of incompleteness.

Permanent employee: Interviewers who are permanent employees of the implementing agency have slightly but significantly lower levels of incompleteness compared with nonpermanent employees. As Table 6.1 shows, nearly half of the cases are for interviewers who were "missing" on this variable.

Summary: Data quality is better for interviewers who are female, are older, have no DHS experience, and are permanent employees. For all covariates, the "missing" category consistently has the lowest prevalence of incompleteness.

Inconsistent ages of women

Figure 6.12a Point estimates and the lower and upper ends of 95% confidence intervals for the prevalence of inconsistency of women's age in the household and individual interviews, within categories of the interviewer's place of residence, age group, sex, and marital status. All surveys are pooled. See Appendix Table A3.4 for numerical values.

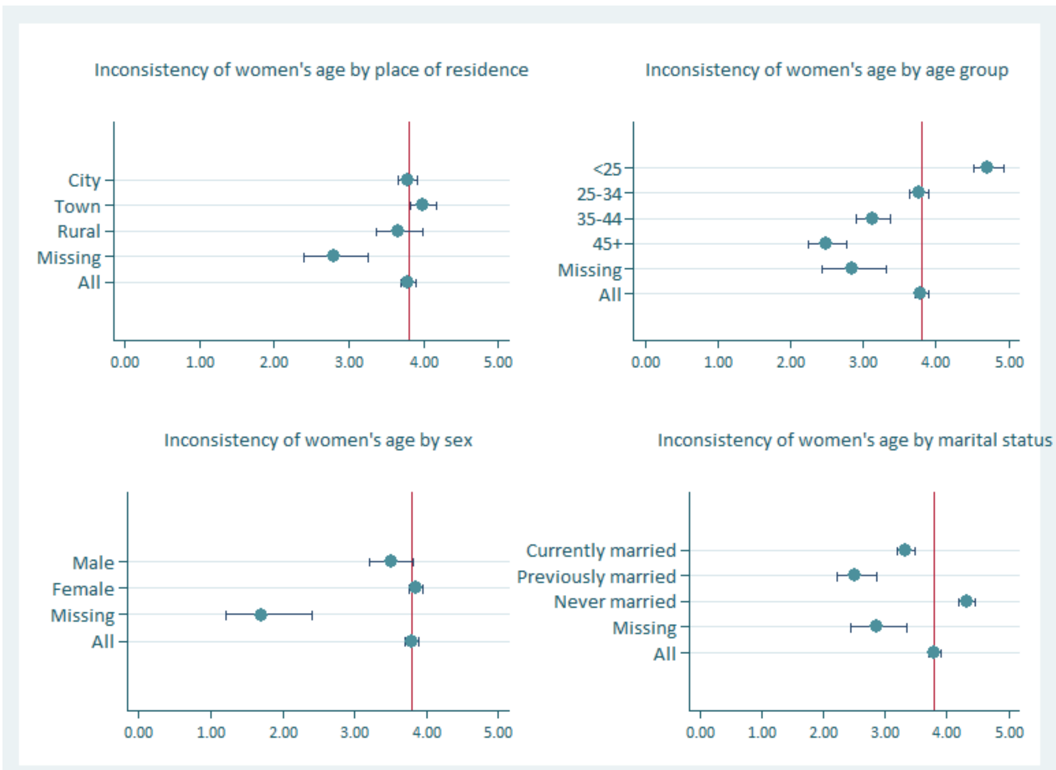
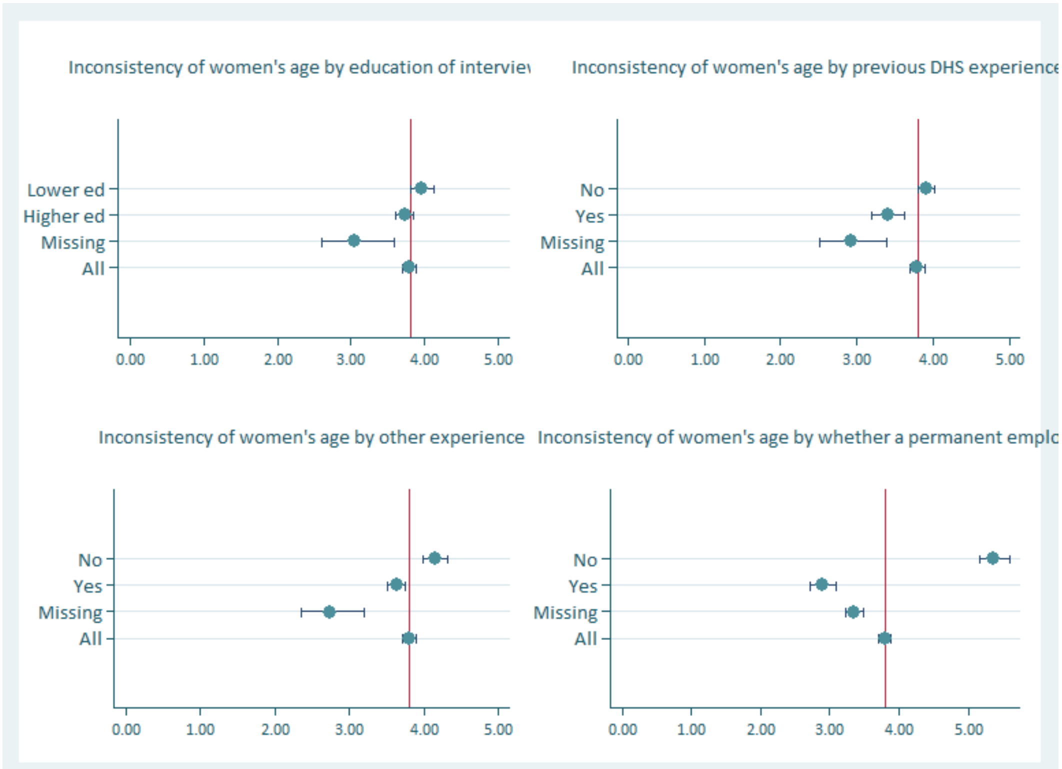


Figure 6.12b Point estimates and the lower and upper ends of 95% confidence intervals for the prevalence of inconsistency of women's age in the household and individual interviews, within categories of the interviewer's education, previous DHS experience, previous experience with other surveys, and whether a permanent employee of the implementing agency. All surveys are pooled. See Appendix Table A3.4 for numerical values.



Place of residence: As with incompleteness, the highest prevalence of this type of inconsistency is in the "town" category, but here the difference from the other categories is not significant.

Age group: There is a monotonic improvement in this indicator with interviewer's increasing age, and the differences between successive age intervals are always highly significant. The range is wide, especially relative to the level. The prevalence is 4.7% in the youngest age group and only 2.5% in the oldest age group.

Sex: Most of the interviewers are women. The difference between women and men is not significant.

Marital status: The "previously married" category is relatively small but has significantly lower prevalence of inconsistency. "Never married" interviewers have significantly higher (worse) prevalence than the other two categories, by 1-2 percentage points.

Education: Interviewers with more education have lower prevalence than those with less education. The difference is small but statistically significant.

Previous DHS experience: Interviewers with previous DHS experience have lower prevalence than those without. The difference is small, about half of a percentage point, but statistically significant.

Other survey experience: Interviewers with other survey experience have lower prevalence than those without. The difference is small, again about half of a percentage point, but statistically significant.

Permanent employee: Interviewers who are permanent employees of the implementing agency have lower prevalence than nonpermanent employees. The difference is substantial—more than 2 percentage points—and highly significant.

Summary: Data quality is better for older interviewers, with a strong pattern of improvement as age increases. Quality improves with more education, previous DHS experience, other survey experience, and permanent employee status. The relationships with the last four covariates are in the desirable direction.

Heaping of ages

Figure 6.13a Point estimates and the lower and upper ends of 95% confidence intervals for the prevalence of heaping of women's age, within categories of the interviewer's place of residence, age group, sex, and marital status. All surveys are pooled. See Appendix Table A3.5 for numerical values.

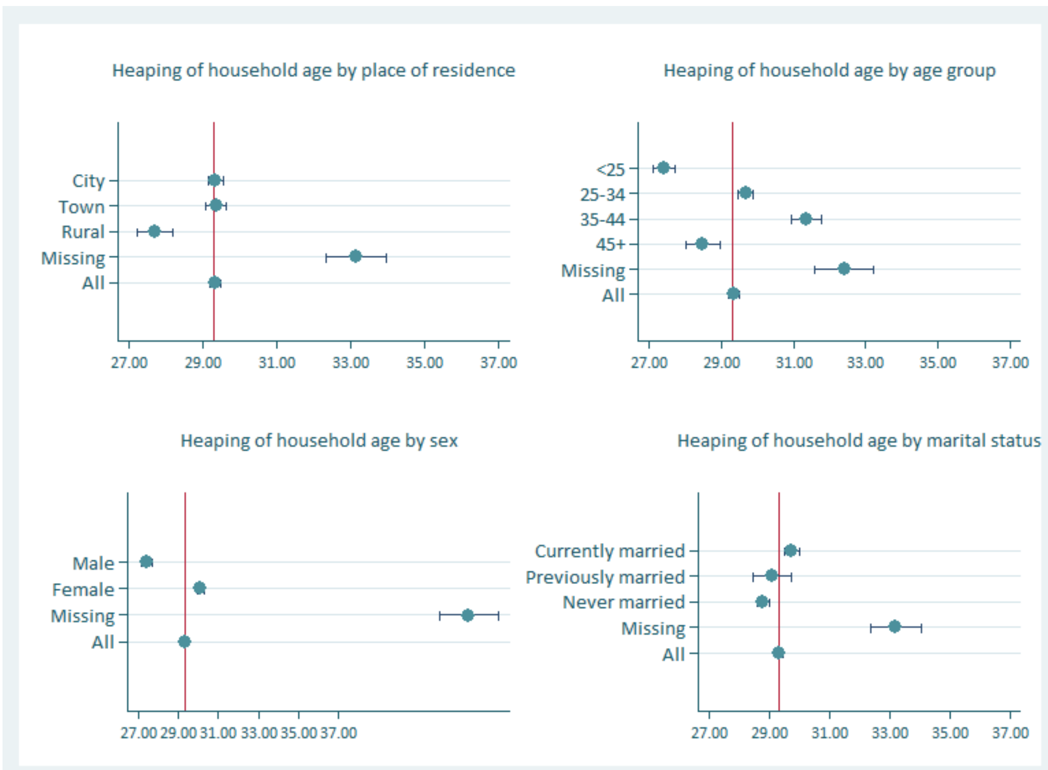
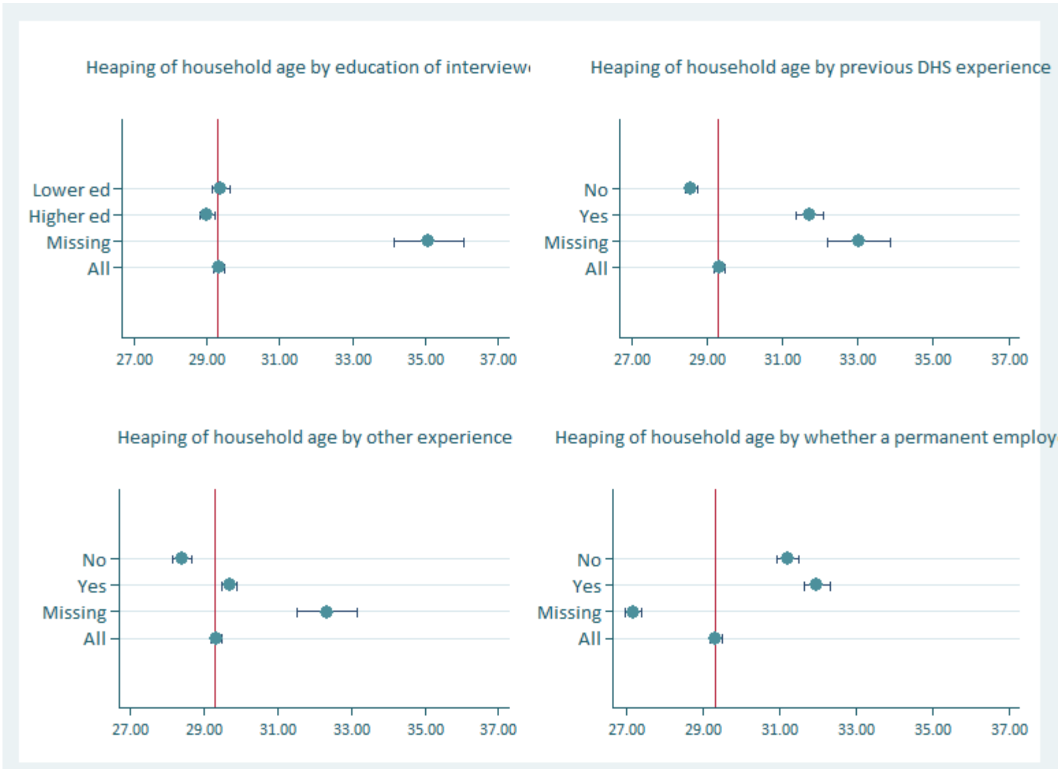


Figure 6.13b Point estimates and the lower and upper ends of 95% confidence intervals for the prevalence of heaping of women's age, within categories of the interviewer's education, previous DHS experience, previous experience with other surveys, and whether a permanent employee of the implementing agency. All surveys are pooled. See Appendix Table A3.5 for numerical values.



Place of residence: Age heaping is worse for "rural" interviewers, by nearly 2 percentage points. There is no difference between "city" and "town".

Age group: The age groups are significantly different from one another but the pattern is difficult to interpret. It is "J-shaped". Heaping is least likely for the youngest age group and then for the oldest age group. It is most likely for the two middle age groups. The range across age groups is about 4 percentage points.

Sex: There is substantially less heaping for male interviewers, by about 3 percentage points.

Marital status: The differences between categories are small. The level of heaping in the "currently married" category is significantly higher than in the "never married" category.

Education: Interviewers with more education have lower levels of heaping than those with less education. The difference is small but statistically significant.

Previous DHS experience: Heaping is substantially and significantly worse among interviewers with prior DHS experience.

Other survey experience: Heaping is also substantially and significantly worse among interviewers with other previous survey experience, although the effect is less than for DHS experience.

Permanent employee: Heaping is substantially and significantly worse for interviewers who are permanent employees, but the effect is small.

Summary: Heaping is strongly related to interviewer's age, but not monotonically. Its relationship to the last three covariates is the opposite of what was seen for age consistency. The residual or "missing" category has the highest level of heaping, also the opposite of what was seen for age consistency. There is evidence of a tradeoff between age consistency and age heaping. If ages are rounded to multiples of 5, then the (rounded) age will tend to be the same in both the household survey and the individual survey, producing a superficial impression of consistency. An interviewer who attempts to be more accurate will avoid rounding, but will be more likely to revise the estimate of age when moving from the household interview to the individual interview.

Displacement of women's ages

Figure 6.14a Point estimates and the lower and upper ends of 95% confidence intervals for the prevalence of downward displacement of women's age in the household survey, within categories of the interviewer's place of residence, age group, sex, and marital status. All surveys are pooled. See Appendix Table A3.6 for numerical values.

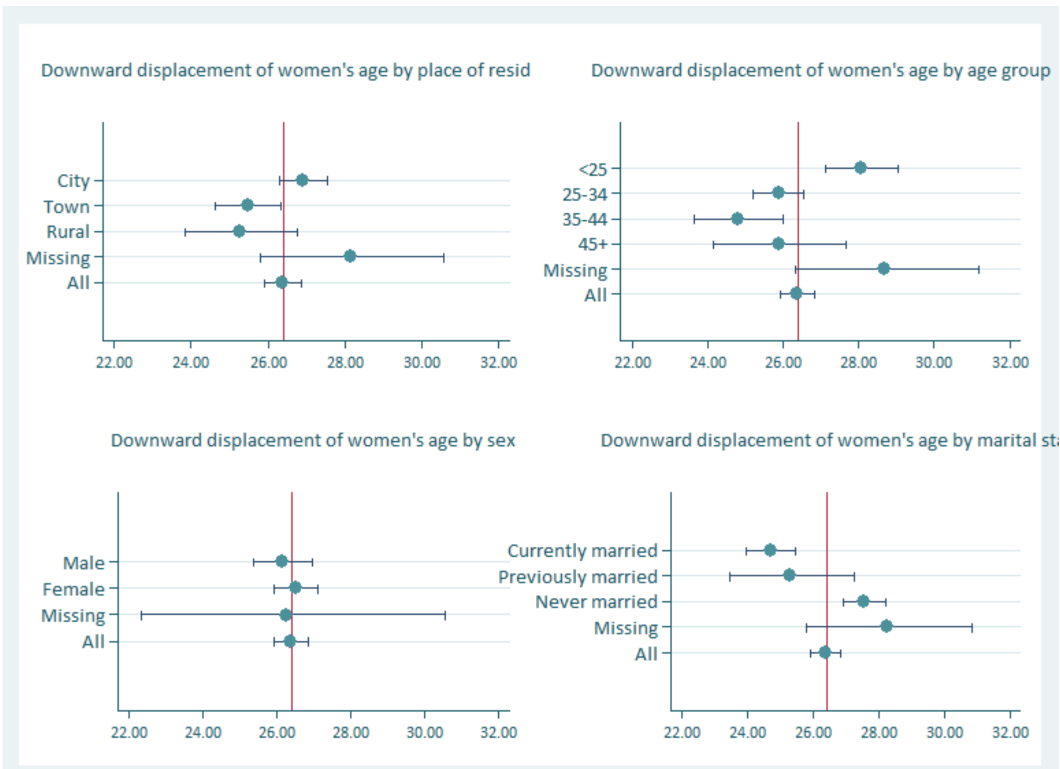
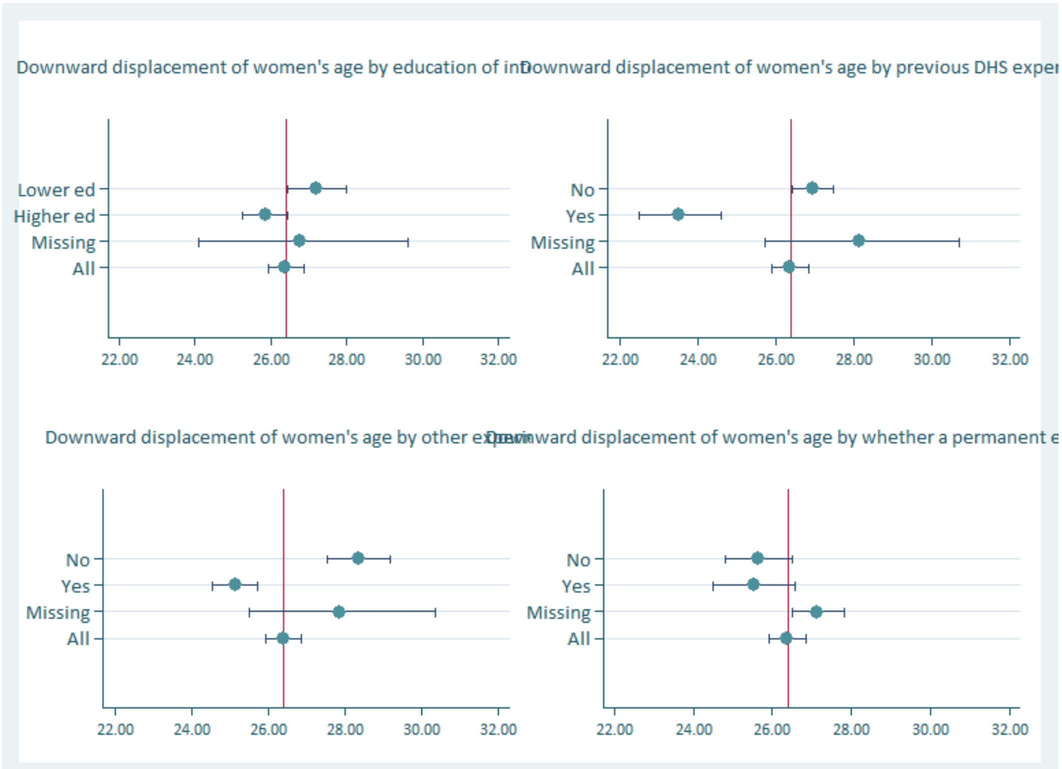


Figure 6.14b **Point estimates and the lower and upper ends of 95% confidence intervals for the prevalence of downward displacement of women's age in the household survey, within categories of the interviewer's education, previous DHS experience, previous experience with other surveys, and whether a permanent employee of the implementing agency.** All surveys are pooled. See Appendix Table A3.6 for numerical values.



Place of residence: Displacement is significantly more prevalent for "city" interviewers than for the other two categories, by more than 1 percentage point.

Age group: Displacement is significantly more prevalent in the youngest age group of interviewers. The other age groups are not significantly different from one another. The range in prevalence is more than 3 percentage points.

Sex: Not significant.

Marital status: Displacement is significantly more prevalent in the never-married category. The other groups are not significantly different from one another. The range in prevalence is about 3 percentage points.

Education: Displacement is significantly more prevalent for interviewers with lower education, by about 1 percentage point.

Previous DHS experience: Displacement is significantly more prevalent for interviewers without DHS experience, by a wide margin of more than 3 percentage points.

Other survey experience: Displacement is also significantly more prevalent for interviewers without other survey experience, by a wide margin of more than 3 percentage points.

Permanent employee: There is no significant difference by employment status. The significance of the variable arises from the "missing" category, which has significantly more displacement than the two employment categories.

Summary: The number of cases for this indicator of misreporting is much smaller than for the previous indicators, so the confidence intervals are wide. More education, previous DHS experience, and other survey experience all have a favorable impact on reducing the level of age displacement. Younger and never-married interviewers, two categories with considerable overlap, have a higher prevalence of displacement.

Over-dispersion of the ages of children under age 5

Figure 6.15a Point estimates and the lower and upper ends of 95% confidence intervals for the prevalence of over-dispersion of children's age, implied by over-dispersion of the HAZ and WAZ scores, within categories of the interviewer's place of residence, age group, sex, and marital status. All surveys are pooled. See Appendix Table A3.7 for numerical values.

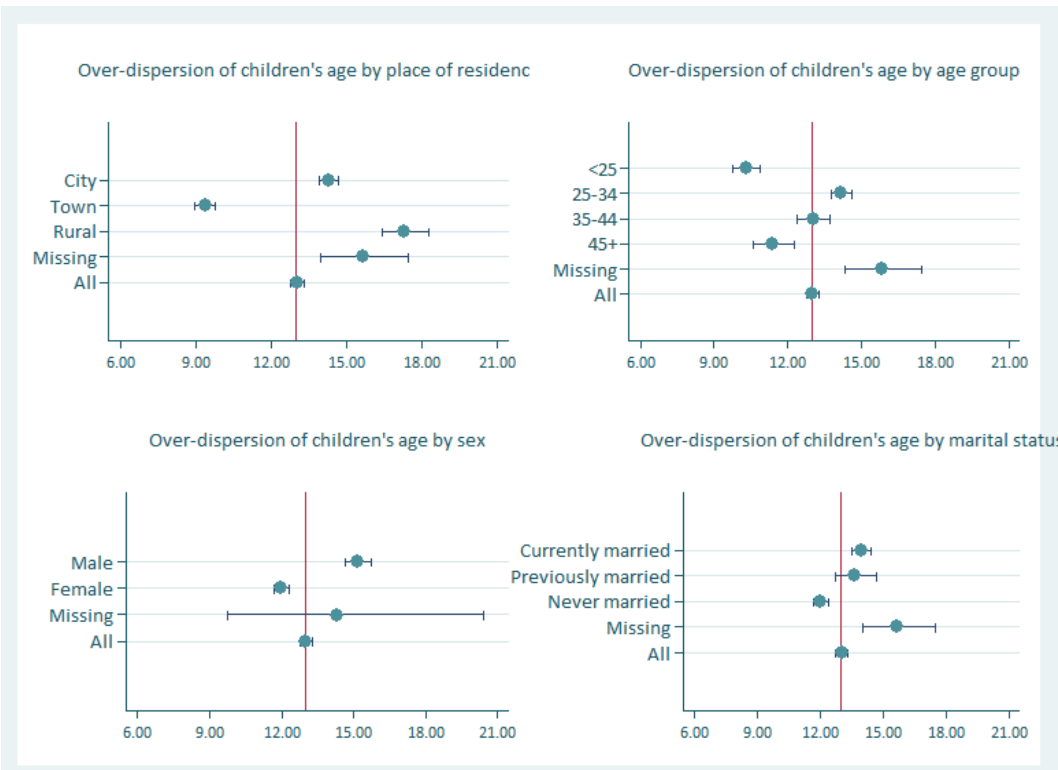
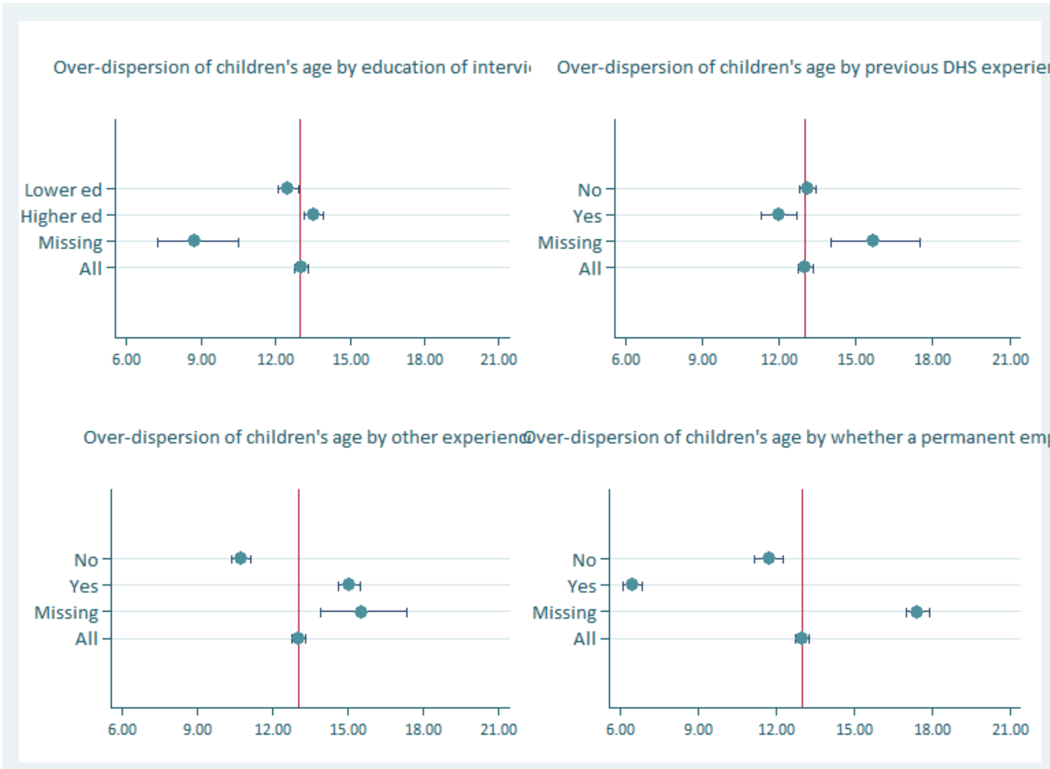


Figure 6.15b Point estimates and the lower and upper ends of 95% confidence intervals for the prevalence of over-dispersion of children's age, implied by over-dispersion of the HAZ and WAZ scores, within categories of the interviewer's education, previous DHS experience, previous experience with other surveys, and whether a permanent employee of the implementing agency. All surveys are pooled. See Appendix Table A3.7 for numerical values.



Place of residence: The lowest prevalence of over-dispersion of children's age is for "town" interviewers, highest is for "rural," and "city" is intermediate. The differences are statistically significant, and the range is large, at 8 percentage points.

Age group: The youngest and oldest interviewer age groups have the lowest prevalence. The middle age groups have significantly higher prevalence. The range is about 4 percentage points.

Sex: The prevalence is about 3 percentage points lower for female interviewers than for male interviewers. The difference is very significant.

Marital status: There is little variation in prevalence by marital status. The prevalence is significantly lower for never-married interviewers than for the other two categories, but the difference is relatively small.

Education: The prevalence is significantly lower for interviewers with less education, but the difference is relatively small, only about 1 percentage point.

Previous DHS experience: The prevalence is significantly lower for interviewers with previous DHS experience, but the difference is only about 1 percentage point.

Other survey experience: By contrast, interviewers with other, non-DHS experience have much higher prevalence of over-dispersion, at about 4 percentage points. In general for most indicators, DHS experience and other survey experience have effects in the same direction. For this indicator, however, the effects are in different directions, with non-DHS experience actually working against better data quality.

Permanent employee: The prevalence is about 5 percentage points less for interviewers who are permanent employees than for those who are not.

Summary: This indicator is exploratory and must be interpreted with caution. Some differentials are worth noting, in particular the substantially lower level of the indicator for female interviewers compared with men, the very low prevalence for permanent employees, and the negative effect of previous experience with non-DHS surveys.

7 DISCUSSION AND CONCLUSIONS

This methodological report has examined variation across interviewers in 25 indicators of data quality in 15 recent DHS surveys.[12] Five indicators relate to nonresponse and refusals; three relate to under-5 mortality; and 17 relate to age and date reporting, which was further divided into five subgroups.

All 25 indicators were defined as binary (0/1) variables for individuals, based on data in the household survey, the survey of women, the birth histories in the survey of women, or the survey of men. For some indicators, "perfect" data would not include any "1" values. For other indicators, even with "perfect" data, "1" would occur about 20% of the time or about 25% of the time or with some other non-zero frequency. For such indicators, the "1" response indicates potential misreporting. For individual respondents, it is not possible to say, for example, that an age ending in 0 or 5 is incorrect, but if some subgroup of respondents has an aggregated prevalence significantly above 20%, then we can infer that there is digit preference or rounding. The levels of these binary indicators, and variations according to characteristics of the respondents, are be analyzed with aggregated logit regressions within the framework of generalized linear models.

There are many potential sources of errors in the surveys. The respondents may not know their true ages, for example, and when asked may tend to give a rounded estimate. It would be expected that such errors are more common for respondents who are rural, older, or less educated, for example, because their birthdates and true ages are often less important in their daily lives. DHS tries during data collection to improve the estimates in settings where the respondents tend not to know ages and dates.

It is reasonable, based on a great deal of related literature on the methodology of survey research, to believe that interviewers are also a potential source of data quality problems, through many possible mechanisms. Most importantly, even the best interviewers can have some motivation to reduce their workload. When confronted with ambiguity about someone's age, if it is near a boundary for eligibility, the interviewer may resolve the ambiguity by putting the person outside the age range of eligibility for the individual interview or for the child-health questions. If an eligible respondent is not immediately available for the individual interview, the interviewer may not make every possible effort to follow up at a later time. These are examples of how the workload may be reduced.

Another example of how an interviewer effect happens is that the interviewer may have genuinely misunderstood some instructions during training and thus under-reports ages ending in 0 or 5. This particular type of misreporting actually requires some effort, because an age that was probably given by the respondent as ending in 0 or 5 must be revised and moved into an adjacent age.

The interviewer may tend to under-report children who died, possibly because of a mistaken belief that they are less important for the research than children who survived. Perhaps the interviewer is uncomfortable with extra prompting for children who died. The interviewer may appear to the respondent to be rushed or

[12] The number of relevant surveys has varied across indicators and according to whether interviewer characteristics were being analyzed. One of the surveys is an AIDS Indicator Survey (AIS) and four are Malaria Indicator Surveys (MIS), but all were conducted through the Demographic and Health Surveys Program.

judgmental or socially distant, leading the respondent to withhold or incorrectly state some information. For every indicator in this report, it is possible to develop at least one plausible mechanism leading to misreporting in which the interviewer plays a role.

The individual-level indicators are aggregated according to the ID codes for the interviewers in order to calculate interviewer-specific levels of prevalence or percentages. These are analyzed with an aggregated version of logit regression that is also within the framework of generalized linear models. This report identifies "interviewer effects" with two criteria. The first criterion, which is most relevant for the indicators that ideally would never take the value "1", is that the overall prevalence must be above some threshold. Every indicator reaches a problematic level in at least some of the surveys. The second criterion is that there must be highly significant variation in prevalence across interviewers.

It is normal in statistical analysis to focus on variation, but we note that there is a risk of overlooking the potential role of the interviewers, or other aspects of survey fieldwork, in contributing to the level of the prevalence of an indicator, not just its variation. For example, suppose that 30% of the ages are rounded to a multiple of 5. The expectation would have been approximately 20%. Statistical models describe the variation around the overall observed prevalence, 30%, rather than around the hypothetical true prevalence of 20%. The justification for 30% as the reference value would be that there are many possible influences on the overall prevalence; it is quite possible that the deviation from 20% is due mostly to the respondents' fundamental uncertainty about their true ages, rather than to any action by the interviewers. But if the interviewers are responsible for some of the variation around the 30%, it would seem plausible that they bear at least some responsibility for the excess over the true level of 20%.

After identifying combinations of indicators and surveys that have substantial levels of misreporting and substantial variation across interviewers, the next step is to describe the degree of concentration of errors and to identify specific interviewers who contribute most to the problematic responses. For many indicators, a handful of specific interviewers are responsible for a large fraction of the total deviance, or variation. It is easy to artificially drop those interviewers and recalculate the prevalence of the indicator and the total deviance. Examples are provided in which both the prevalence and the deviance change dramatically in such a simulation. It probably would have been possible to identify these problematic interviewers in real time, during data collection, and intervene to improve their reporting. DHS has always tried to identify data collection problems with field check tables, and that approach is being upgraded to dashboards. The current approach leads to many interventions during fieldwork, but it is clearly not foolproof.

Some indicators, such as incompleteness in reporting women's ages—usually taking the form of a missing month of birth—appear with both very high frequencies and very high variation across interviewers. For these combinations of indicators and surveys, it is not be possible to identify just a few interviewers who are contributing disproportionately to the prevalence of misreporting, and without whom the variation would not be significant. In some such instances it appears to be the supervisors who are problematic, and all the interviewers on their team are being monitored too loosely or too strictly. If broadly dispersed variability could be identified during fieldwork, then retraining of supervisors or new across-the board guidelines would be justified.

The analysis of interviewer effects, as described above, only requires the consistent use of interviewer ID codes. Many older surveys could be examined in the same way, although it is believed that in the early rounds of DHS, two or more interviewers would often share the same ID code, and the ID codes do not clearly identify teams. Sharing of codes or other inconsistencies would attenuate effects but not completely subvert an analysis of trends and differences. DHS household surveys now routinely include a survey of the interviewers, and one motivation for this report has been to determine whether the characteristics in the interviewer survey are systematically related to the quality of data in the DHS survey. Chapter 2 of this report presents a thorough description of the interviewer survey, the composition of the interviewers in the 14 countries, and some of the different patterns found in the fieldwork, apart from the relationship to the prevalence of the data quality indicators.

In the analysis of the bivariate association between misreporting indicators and interviewer characteristics or covariates, the data were pooled across surveys and a simple glm model was estimated without any adjustment for the differences between the surveys. In Chapter 4, all five indicators of nonresponse and refusal were pooled. In Chapter 5, the two indicators of potential omission of early infant and under-5 deaths were pooled. In Chapter 6, five of the 17 indicators were used alone (without pooling of indicators), one for each of the five types of age and date misreporting.

Nearly every bivariate association between an indicator and a covariate was statistically significant at a high level. For most indicators, the pattern of association across categories was interpretable. We will not review the patterns here but will mention that, for the most part, higher age and higher level of education were beneficial for data quality. The effect of previous experience with a DHS survey and/or another survey was inconsistent, and not always beneficial. For example, age heaping appears to be worse for experienced interviewers.

Perhaps the most important limitation of this analysis is that the interviewers are not assigned randomly to several different geographic areas of a country, according to an experimental design. The assignments to households within clusters are random, and the assignment to clusters within districts, and possibly within regions,[13] are intended to be random, but a specific interviewer generally works entirely within a specific region. As a result, the effects of interviewers (and teams as a whole) are confounded with regional variation. Many socioeconomic characteristics of the sample households and the respondents tend to be spatially structured—to be more consistent within regions than across regions. The urban-rural composition varies a great deal across regions. The region that includes the country's capital city is often much more developed than the hinterland. We expect that some of the variation in data quality that has been attributed to interviewers could be traced to the composition or characteristics of the different subnational areas. In pooling surveys, some variation is likely due to differences between countries in composition and characteristics.

This analysis has not included characteristics of the respondents. It is likely that those characteristics are also systematically related to the indicators of data quality, and it would be helpful to control for those effects as part of assessing the interviewer effects (or, from an alternative perspective, to control for interviewer effects as part of assessing respondent effects). There would be particular interest in interactions between these two kinds of effects, as suggested earlier. The role of social distance, indicated by differences between interviewer and respondent in age, sex, residence, and education, could shed light on how the

[13] In DHS terminology, regions are the level 1 administrative units, and districts are level 2.

interviewer's ability to establish confidence and rapport with the respondent may be beneficial for data quality.

It was initially intended that this report would include analysis of respondent characteristics and measures of social distance between respondent and interviewer. Apart from size limitations, a major reason for not proceeding in that direction is that such analysis should be survey-specific, rather than pooled across countries. Apart from specific examples, most of this report uses an analytical strategy of comparing countries or pooling countries. Country-specific analysis would require a different structure. Further research is planned that will consist of a small number of case studies. Research is also underway to explore interviewer effects with sensitive questions related to sexual activity, for which interviewer/respondent differences and interactions are relevant.

It would be desirable in future methodological analyses to examine the effects of interviewers—or fieldworkers more generally—on other types of data in DHS surveys. In addition to sensitive questions related to sexual activity, topics could include anthropometric measurement, potential use of traditional methods of family planning, the completeness of data in the contraceptive calendar, the sibling histories, the domestic violence module, and the child discipline module. Any question that includes a "don't know" category, or variable that has missing values (distinct from Not Applicable), may potentially identify room for improvement by interviewers.

This report has demonstrated a strategy involving the specification of individual-level binary indicators, the identification of problematic combinations of surveys and indicators, the identification of problematic interviewers or groups of interviewers, the selection of diagnostic indicators within groups of indicators, and the application of generalized linear models to identify significant covariates. This strategy can be broadly applied to any type of error or potential error, beyond the types explicitly included here.

Even more important than further research to identify and interpret interviewer effects, the findings in this report and other related reports should be used to improve the quality of future surveys. Many important indicators of data quality show excellent levels in most DHS surveys. For example, nonresponse is very low. Problems with age and date transfers, particularly the transfer of children's birth dates across the boundary for the child health questions, have declined. Incompleteness of age has steadily declined and is currently limited almost exclusively to a missing month of birth (Pullum and Staveteig 2017). Nevertheless, some problems, such as omission of children who died, are encountered repeatedly in successive surveys in certain countries. It can be expected that some problems, such as nonresponse and refusals, will become more prevalent as the standard of living improves in some countries; urban and better-educated areas have seen increases in their rates of nonresponse and refusal.

In order to maintain and further improve the quality of DHS surveys, problematic interviewers should be identified during fieldwork, not afterwards. They can then be retrained or monitored more closely. If a problem appears to be at the level of a team, or a higher organizational level—as when several teams are overcorrecting for heaping or displacement and others are not correcting at all—interventions can take place during fieldwork. In addition, the field check tables that are currently prepared during fieldwork in all surveys should be reconfigured with binary indicators that can be analyzed with statistical software rather than having to be scanned visually.[14] Future fieldwork should be accompanied with real-time monitoring of data quality indicators, disaggregated as much as possible. Finally, at the completion of every survey a standard assessment of data quality should be prepared, made publicly available, and archived, going far beyond the five tables that currently comprise Appendix C of the main country report. Movement in the direction of better data is the fundamental goal of this report.

[14] A similar purpose is served by a fieldwork "dashboard" that DHS is currently developing.

REFERENCES

Adida, C. L., K. E. Ferree, D. N. Posner, and A. L. Robinson. 2016. “Who’s Asking? Interviewer Coethnicity Effects in African Survey Data.” *Comparative Political Studies* 49 (12): 1630-1660. https://doi.org/10.1177%2F0010414016633487.

Anglewicz, P. 2009. “The Malawi Diffusion and Ideational Change Project 2004–06: Data Collection, Data Quality, and Analysis of Attrition.” *Demographic Research* 20 (21): 503. https://dx.doi.org/10.4054/DemRes.2009.20.21.

Arnold, F. 1991. “An Assessment of Data Quality in the Demographic and Health Surveys.” *Proceedings of the Demographic and Health Surveys World Conference*, pp. 785-798. Columbia, Maryland, USA: IRC/Macro International.

Bignami-Van Assche, S., G. Reniers, and A. A. Weinreb. 2003. “An Assessment of the KDICP and MDICP Data Quality: Interviewer Effects, Question Reliability and Sample Attrition. *Demographic Research* 1 (2): 31-76. https://dx.doi.org/10.4054/DemRes.2003.S1.2.

Blanc, A. K., and N. Rutenberg. 1990. “Assessment of the Quality of Data on Age at First Sexual Intercourse Age at First Marriage and Age at First Birth in the Demographic and Health Surveys.” In: *An Assessment of DHS-I Data Quality*. DHS Methodological Reports No. 1. Columbia, Maryland, USA: Institute for Resource Development / Macro Systems, Inc. https://www.dhsprogram.com/pubs/pdf/MR1/MR1.pdf

Blom A., and J. M. Kornbmacher. 2013. “Measuring Interviewer Characteristics Pertinent to Social Surveys: A Conceptual Framework.” *Survey Methods: Insights from the Field.* Retrieved from https://surveyinsights.org/?p=817

Cleary, P. D., D. Mechanic, and N. Weiss. 1981. “The Effect of Interviewer Characteristics on Responses to a Mental Health Interview.” *Journal of Health and Social Behavior* 2 (2): 183-193. https://www.jstor.org/stable/2136294.

Croft, T. N. 1991. *DHS Data Editing and Imputation*. Calverton, Maryland, USA: Demographic and Health Surveys. 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. http://dhsprogram.com/pubs/pdf/OP3/OP3.pdf.

Davis, R. E., M. P. Couper, N. K. Janz, C. H. Caldwell, and K. Resnicow. 2009. “Interviewer Effects in Public Health Surveys." *Health Education Research* 25 (1): 14-26. https://dx.doi.org/10.1093%2Fher%2Fcyp046.

Espeut, D., and S. Becker. 2015. "The Validity of Birth and Pregnancy Histories in Rural Bangladesh." *Journal of Health, Population and Nutrition* 33 (1): 17. https://doi.org/10.1186/s41043-015-0027-8.

Ford, K., and A. E. Norris. 1997. "Effects of Interviewer Age on Reporting of Sexual and Reproductive Behavior of Hispanic and African American Youth." *Hispanic Journal of Behavioral Sciences* 19 (3): 369-376. https://doi.org/10.1177%2F07399863970193008.

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. http://dhsprogram.com/pubs/pdf/OP4/OP4.pdf.

Hill, K., and Y. Choi. 2006. "Neonatal Mortality in the Developing World." *Demographic Research* 14 (18): 429-52. https://dx.doi.org/10.4054/DemRes.2006.14.18.

Himelein, K. 2015. "Interviewer Effects in Subjective Survey Questions: Evidence from Timor-Leste." *International Journal of Public Opinion Research* 28 (4): 511-533. DOI: 10.1093/ijpor/edv031.

Howell, N. 1986. "Age Estimates and Their Evaluation in Research." In: *New Methods for Old-age Research: Strategies for Studying Diversity.* [Edited by] Christine L. Fry and Jennie Keith, pp. 57-76. South Hadley, Massachusetts, USA: Bergin and Garvey Publishers.

Hox, J. J., E. D. De Leeuw, and G. Kreft. 1991. "The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data: A Multilevel Model." In: Biemer, Groves, Lyberg, Mathiowetz, and Sudman (eds.): *Measurement Errors in Surveys*. New York, New York, USA: John Wiley & Sons, Inc.

ICF. 2017. *Demographic and Health Survey Interviewer's Manual*. Rockville, Maryland, USA: ICF. https://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. http://dhsprogram.com/pubs/pdf/DHSQ6/DHS6_Questionnaires_5Nov2012_DHSQ6.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. http://dhsprogram.com/pubs/pdf/AS19/AS19.pdf.

Liu, M. 2016. "The Effect of the Perception of an Interviewer's Race on Survey Responses in a Sample of Asian Americans." *Asian American Journal of Psychology* 7 (3): 167.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd edition, London, England: Chapman and Hall.

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

Neal, S., C. Ruktanonchai, V. Chandra-Mouli, Z. Matthews, and A. J. Tatem. 2016. "Mapping Adolescent First Births within Three East African Countries Using Data from Demographic and Health Surveys:

Exploring Geospatial Methods to Inform Policy." *Reproductive Health* 13 (1): 98. https://dx.doi.org/10.1186%2F1742-4755-12-13.

Nix, B. T. 2014. "The Effect of Interview Length on Data Quality in the Consumer Expenditure Interview Survey." *Proceedings of the Joint Statistical Meetings of the American Statistical Association, Government Statistics Section*, pp. 1405-1414.

ORC Macro. 2001a. *Model "A" Questionnaire with Commentary for High Contraceptive Prevalence Countries*. ORC Macro. 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*. Calverton, Maryland, USA: ORC Macro. 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. 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. http://dhsprogram.com/pubs/pdf/MR11/MR11.pdf

Pullum, T. W., and S. 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. http://dhsprogram.com/pubs/pdf/MR19/MR19.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. 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. http://dhsprogram.com/pubs/pdf/MR12/MR12.pdf.

Wang, K., P. Kott, A. Moore, T. G. Virag, and S. Abuse. 2013. *Assessing the Relationship between Interviewer Effects and Nsduh Data Quality*. Report prepared by Research Triangle Institute for the Substance Abuse and Mental Health Services Administration, Research Triangle Park, NC, USA.

Yang, M.-L., and R.-R. Yu. 2008. "The Interviewer Effect when There Is an Education Gap with the Respondent: Evidence from a Survey on Biotechnology in Taiwan." *Social Science Research* 37 (4): 1321-1331.

APPENDIX 1 TABLES TO ACCOMPANY CHAPTER 2

Appendix Table A1.1 Percentage distribution of interviewers by their ethnicity

DHS		MIS					
	%		%		%		%
Armenia	N/A	**Ghana**		**Nigeria (n=195)**		**Sierra Leone**	
Burundi	N/A	Akan	45.9	Igbo/Ibo	17.4	Mande	31.9
Malawi	N/A	Ewe	23.0	Yoruba	17.4	Temne	16.8
Myanmar	N/A	Ga/Dangme	13.1	Hausa	11.3	Limba	11.5
Nepal (n=91)		Mole-Dagbani	6.6	Missing	5.1	Krio	8.9
Hill Brahmin	34.1	Guan	4.9	Igala	3.6	Madingo	6.2
Newar	18.7	Grusi	4.9	Ijaw/Izon	3.1	Fullah	4.4
Hill Janajati	17.6	Other	1.6	Ibibio	2.1	Susu	3.5
Hill Chhetri	14.3	Gurma	0.0	Nupe	2.1	Loko	2.7
Other Terai caste	8.8	Mande	0.0	Fulani	1.5	Kissi	0.9
Terai							
Brahmin/Chhetri	2.2	**Kenya**		Kanuri/Beriberi	1.5	Kono	0.9
Hill Dalit	2.2	Kikuyu	19.6	Tiv	1.5	Koranko	0.9
Terai Janajati	1.1	Kalenjin	12.8	Afo	1.0	Missing	11.5
Muslim	1.1	Luo	12.8	Bini/Edo	1.0		
Uganda		Kamba	9.0	Delta Ibo	1.0	**AIS**	
Baganda	20.5	Luhya	7.5	Eggon	1.0	**Mozambique**	N/A
Banyankore	8.6	Kisii	6.8	Higgi	1.0		
Iteso	8.0	Somali	6.8	Ichen	1.0		
Bagisu	7.3	Other	6.0	Idoma	1.0		
Basoga	6.6	Embu	3.0	Igede	1.0		
Batoro	6.6	Mijikenda	3.0	Owo	1.0		
Lango	6.0	Borana	3.0	Tangale	1.0		
Bakiga	4.6	Turkana	3.0	Urhobo	1.0		
Banyoro	4.6	Maasai	1.5	Gwari	1.0		
Lugbara	4.6	Meru	1.5	Others	1.0		
Acholi	2.7	Taita/Taveta	1.5	Abua	0.5		
Alur	2.7	Swahili	1.5	Agbassa	0.5		
Basamia	2.0	Pokot	0.8	Alago	0.5		
Jonam	2.0	**Liberia**	N/A	Annang	0.5		
Bafumbira	1.3	Chewa	22.6	Arab	0.5		
Bagwere	1.3	Lomwe	20.3	Arugu	0.5		
Bahororo	1.3	Ngoni	15.2	Auchi	0.5		
Ethur	1.3	Tumbuka	13.2	Aulliminden	0.5		
Kumam	1.3	Yao	7.4	Ayetoro Gbede	0.5		
Aringa	0.7	Other	4.7	Badarare	0.5		
Bakonzo	0.7	Tonga	4.4	Bajju	0.5		
Batagwenda	0.7	Sena	4.1	Bolawa/Bolewa	0.5		
Jopadhola	0.7	Nkhonde	1.7	Bura/Babur	0.5		
Missing	4.0	Missing	6.4	Degima	0.5		
Zimbabwe	N/A			Ebira/Igbira	0.5		
				Efik	0.5		
				Esan	0.5		
				Gamaru	0.5		
				Gbaju/Gbagi	0.5		
				Guemai	0.5		
				Isoko	0.5		
				Kagoma	0.5		
				Kataf/Atyap	0.5		
				Kuteb	0.5		
				Mada	0.5		
				Marghi/Mangi	0.5		
				Mwagavol	0.5		
				Ndola	0.5		
				Ngoshe	0.5		
				Ogoja	0.5		
				Owan	0.5		
				Tarok	0.5		
				Wari	0.5		
				Wodaabe	0.5		
				Yakurr	0.5		
				Bakabe	0.5		
				Cham	0.5		
				Gaanda	0.5		
				Don't know	0.5		

N/A = Not available

Appendix Table A1.2 Percentage distribution of interviewers by their religion

DHS	%	MIS	%
Armenia	N/A	**Ghana**	
Burundi		Pentecostal/Charismatic	41.0
Catholique	72.8	Presbyterian	16.4
Protestante	22.8	Catholic	14.8
Adventiste	2.5	Methodist	11.5
Autre	1.3	Islam	6.6
Temoin de Jehovah	0.6	Anglican	4.9
Malawi		Other Christian	4.9
Ccap	34.1	**Kenya**	
Catholic	18.9	Roman Catholic	23.3
Christian	17.9	Protestant/other Christian	59.4
7th day Adventist/Baptist	12.8	Muslim	14.3
Anglican	2.7	No religion	2.3
Muslim	2.4	Other	0.8
Other	4.7	**Liberia**	
Missing	6.4	Christian	98.3
Myanmar	N/A	Muslim	1.7
Nepal		**Nigeria**	
Hindu	89.0	Christianity	74.9
Buddhist	5.5	Islam	25.1
Kirat	4.4	**Sierra Leone**	
Muslim	1.1	Christian	62.8
Uganda		Muslim	25.7
Anglican	37.8	Missing	11.5
Catholic	31.8		
Pentecostal/Born Again/Evangelical	21.9	**AIS**	
Muslim	1.3	**Mozambique**	
Seventh Day Adventist	2.0	Católica	52.0
Other	1.3	Islâmica	15.1
Missing	4.0	Protestante	13.4
Zimbabwe		Evangélica/pentecostal	11.7
Pentecostal	37.5	Zione/Sião	0.6
Protestant	28.3	Anglicana	2.2
Roman Catholic	19.2	Sem religião	1.7
Other Christian	6.7	Outra	3.4
Other	1.7		
Traditional	0.8		
Apostolic sect	0.8		
Missing	5.0		

N/A = Not available

Appendix Table A1.3 Percentage distribution of interviewers by their mother tongue

DHS		MIS			
	%		%		%
Armenia		**Ghana**		**Sierra Leone**	
Armenian	100.0	Akan	47.5	Mende	32.7
Burundi		Ga	16.4	Krio	16.8
Kirundi	100.0	Ewe	21.3	Temne	12.4
Malawi		Other	14.8	English	9.7
Chichewa	68.9	**Kenya**		Limba	5.3
Tumbuka	14.5	Embu	3.0	Sherbro	3.5
Other	10.1	English	1.5	Loko	2.7
Missing	6.4	Kalenjin	12.0	Fullah	1.8
Myanmar		Kamba	8.3	Madingo	0.9
Myanmar	55.1	Kikuyu	19.6	Kono	0.9
Kayin (Karen)	9.5	Kisii	6.8	Susu	0.9
Rakhine	7.5	Luhya	6.8	Koranko	0.9
Chin	4.8	Luo	12.0	Missing	11.5
Shan	3.4	Maasai	1.5	**AIS**	
Mon	0.7	Meru	1.5	**Mozambique**	30.7
Other	19.1	Mijikenda	3.0	Portuguese	19.0
Nepal		Somali	6.0	Emakhuwa	12.9
Nepali	71.4	Taita/Taveta	1.5	Xichangana	8.9
Newari	8.8	Borana	6.0	Outra	7.3
Maithili	7.7	Pokot	0.8	Cinyungwe	5.6
Tharu	2.2	Turkana	3.0	Echuwabo	5.0
Other	9.9	Swahili	3.0	Cisena	2.8
Uganda		Other	3.8	Xitswa	2.8
Luganda	24.5	**Liberia**		Shona	2.2
Runyankole/Rutero	15.9	Bassa	13.3	Cindau	1.1
Luo	12.6	Gio	5.0	Elomwe	1.1
Ateso	8.6	Gola	6.7	Ciyao	0.6
Lusoga	6.6	Grebo	11.7	Cinyanja	0.0
Lugbara	5.3	Kissi	3.3		
Missing	4.0	Kpelle	13.3		
English	0.7	Krahn	1.7		
Ngakarimojong	0.7	Kru	3.3		
Other	21.2	Lorma	3.3		
Zimbabwe		Mano	15.0		
Shona	78.3	Vai	5.0		
Ndebele	11.7	English	18.3		
Namibia	2.5	**Nigeria**			
English	1.7	Other	31.3		
Venda	0.8	Hausa	22.6		
Tonga	0.8	Igbo	21.0		
Missing	4.2	Yoruba	20.5		
		English	4.1		
		Missing	0.5		

APPENDIX 2 GLM MODELS AND RESIDUAL DEVIANCE

As described in the text, an individual-level outcome Y is defined to have the value 1 if there is a problem or potential problem, such as nonresponse or an age ending in 0 or 5, and the value 0 otherwise, except that the outcome is NA (coded with a dot or period in Stata) if not applicable. This appendix focusses on a single generic outcome in a single country.

In this report, in chapters 4-6, characteristics of the respondents are not used, because the interest is in variations across interviewers and, potentially, according to characteristics of the interviewers. Future analysis may extend to include characteristics of the respondents and how they may interact with characteristics of the interviewers. Because of this simplification for chapters 4-6, the outcome can be reduced to n1 and n0, the number of times the outcome was 1 or 0, respectively, for each interviewer, using "collapse (sum)" in Stata. We then calculate n=n1+n0. This is a file in which interviewers are the units of analysis, and each of them has values of n1, n0, and n. The overall prevalence is P1=N1/N, where N1 and N are the totals of n1 and n, respectively, across all interviewers.

The glm command for a logit regression with aggregated data is "glm n1, family(binomial n) link(logit)". The output from this command will include a single coefficient, b0, and the total deviance, D. The overall prevalence will match with the antilogit of b0, i.e., P1=exp(b0)/[1+exp(b0)]. If there are K interviewers, then D has a chi-square distribution with K-1 degrees of freedom; the output will include the degrees of freedom and the p-value for the chi-square test of the null hypothesis that all interviewers have the same probability of the response.

The post-estimation command to get Stata's deviance residuals, described as "dr" in the text, is "predict dr, deviance". Then "dr2" is obtained with "gen dr2=dr^2". It can be confirmed that the sum of the squared values, obtained with "total dr2" is exactly the same as the total deviance for the model, D.

Further define P0=1-P1 and the interviewer-specific prevalences p1=n1/n (and p0=1-p1). It can be confirmed that the interviewer-specific dr2 terms are exactly the same as the deviance residuals defined by McCullagh and Nelder (1989) as $2[n_0 \log\left(\frac{p_0}{P_0}\right) + n_1 \log\left(\frac{p_1}{P_1}\right)]$. These terms are all positive and have chi-square distributions with one degree of freedom. The sum of the terms across all interviewers is the deviance D.

This data setup is very similar to a table in which there are K rows, one for each interviewer, and two columns labelled "Y=1" and "Y=0", and the cells of the table consist of the frequencies n1 and n0 in each row. A possible test of homogeneity of interviewers would use the familiar Pearson chi-square statistic, which is very close in value to the total deviance, D. (In this context D would also be known as the LR, or likelihood ratio, chi-square.) The sum of the two Pearson chi-square terms in a row of this table would match almost exactly with row-specific values of dr2. The square root of those row totals, if assigned a negative value when p1 is less than P1, will match almost exactly with dr. This correspondence is noted because most readers will be familiar with the test of homogeneity in a two-way table using the Pearson chi-square.

A slightly different approach is required if the data have not been collapsed and the units of analysis are respondents, which will be the data setup if the characteristics of respondents are to be included in the analysis. The outcome is Y, which takes the values 1 or 0. First, the interviewer ID code must be converted to a complete set of dummy or binary variables, with "xi, noomit i.interviewer_id" and "rename _I* *". Second, the logit regression would take the form "logit Y inter*, nocons". Under this specification of fixed effects, there will not be a reference category or reference interviewer. For each interviewer there will be a coefficient that is the log of the odds for that interviewer and a z statistic that will match with dr. The square of z for each interviewer will be dr2, and the sum of the squares will be the deviance D for the model.

APPENDIX 3 TABLES FOR COVARIATE MODELS

Appendix Table A3.1 Estimates and significance tests for the relationship of the interviewer characteristics to the prevalence of nonresponse and refusals. Five indicators and all surveys are pooled.

Covariate	Category	Prevalence	Lower bound	Upper bound	N	Percent	LRchi2	p-value
Residence	City	3.4	3.4	3.5	204,744	57.6	79.8	<0.001
	Town	3.4	3.3	3.5	101,524	28.6		
	Rural	2.6	2.4	2.7	35,849	10.1		
	Missing	3.1	2.8	3.4	13,231	3.7		
Age group	<25	3.8	3.7	3.9	77,506	21.8	94.2	<0.001
	25-34	3.2	3.2	3.3	181,786	51.2		
	35-44	3.0	2.9	3.2	52,061	14.6		
	45+	2.9	2.7	3.1	29,576	8.3		
	Missing	3.7	3.4	4.0	14,419	4.1		
Sex	Male	3.9	3.8	4.0	146,151	41.1	314.8	<0.001
	Female	2.9	2.8	3.0	206,466	58.1		
	Missing	1.1	0.8	1.6	2,731	0.8		
Marital status	Currently married	3.4	3.3	3.5	133,895	37.7	5.3	0.153
	Previously married	3.4	3.2	3.7	19,590	5.5		
	Never married	3.3	3.2	3.4	189,676	53.4		
	Missing	3.0	2.7	3.3	12,187	3.4		
Education	Less educated	4.2	4.1	4.3	155,502	43.8	632.9	<0.001
	More educated	2.6	2.6	2.7	189,837	53.4		
	Missing	3.0	2.7	3.4	10,009	2.8		
DHS experience	No	3.4	3.3	3.5	287,886	81.0	25.7	<0.001
	Yes	3.0	2.8	3.1	55,028	15.5		
	Missing	3.1	2.8	3.4	12,434	3.5		
Other experience	No	3.8	3.7	3.9	141,261	39.8	147.1	<0.001
	Yes	3.0	2.9	3.1	200,940	56.5		
	Missing	3.3	3.0	3.7	13,147	3.7		
Permanent employee	No	3.0	2.8	3.1	85,571	24.1	137.1	<0.001
	Yes	2.8	2.7	3.0	62,418	17.6		
	Missing	3.6	3.5	3.7	207,359	58.3		
All	All	3.3	3.3	3.4	355,348	100.0		

Appendix Table A3.2 Estimates and significance tests for the relationship of the interviewer characteristics to the prevalence of early infant and under-5 deaths. Two indicators and all surveys are pooled.

Covariate	Category	Prevalence	Lower bound	Upper bound	N	Percent	LRchi2	p-value
Residence	City	62.3	61.7	62.9	25,587	56.2	36.6	<0.001
	Town	61.0	60.1	61.8	13,266	29.1		
	Rural	57.9	56.5	59.3	4,956	10.9		
	Missing	59.9	57.6	62.2	1,719	3.8		
Age group	<25	60.8	60.0	61.7	13,200	29.0	22.7	<0.001
	25-34	61.9	61.3	62.6	22,011	48.3		
	35-44	59.3	58.0	60.6	5,601	12.3		
	45+	63.8	62.0	65.4	2,990	6.6		
	Missing	60.0	57.7	62.3	1,726	3.8		
Sex	Male	63.0	60.7	65.3	1,683	3.7	2.1	0.348
	Female	61.3	60.8	61.7	43,611	95.8		
	Missing	60.3	53.8	66.3	234	0.5		
Marital status	Currently married	60.3	59.6	61.1	17,326	38.1	14.9	0.002
	Previously married	62.4	60.6	64.2	2,774	6.1		
	Never married	62.0	61.4	62.7	23,709	52.1		
	Missing	59.9	57.6	62.2	1,719	3.8		
Has children	No	62.5	61.9	63.0	27,067	59.5	35.4	<0.001
	Yes	59.7	58.9	60.4	16,820	36.9		
	Missing	60.0	57.6	62.4	1,641	3.6		
Had a child death	No	61.4	60.9	61.8	42,574	93.5	2.5	0.289
	Yes	62.7	59.9	65.4	1,192	2.6		
	Missing	59.9	57.6	62.1	1,762	3.9		
Education	Less educated	59.9	59.2	60.6	17,562	38.6	31.5	<0.001
	More educated	62.4	61.8	63.0	26,468	58.1		
	Missing	59.2	56.6	61.6	1,498	3.3		
DHS experience	No	61.7	61.3	62.2	37,013	81.3	13.9	0.001
	Yes	59.5	58.3	60.6	6,774	14.9		
	Missing	59.9	57.6	62.2	1,741	3.8		
Other experience	No	61.8	61.1	62.5	17,837	39.2	3.4	0.182
	Yes	61.1	60.5	61.7	25,935	57.0		
	Missing	60.1	57.8	62.4	1,756	3.9		
Permanent employee	No	61.9	61.0	62.7	13,423	29.5	4.1	0.128
	Yes	61.7	60.7	62.7	8,565	18.8		
	Missing	60.9	60.3	61.5	23,540	51.7		
All	All	61.3	60.9	61.8	45,528	100.0		

Appendix Table A3.3 **Estimates and significance tests for the relationship of the interviewer characteristics to the prevalence of incomplete age in the survey of women.** All surveys are pooled.

Covariate	Category	Prevalence	Lower bound	Upper bound	N	Percent	LRchi2	p-value
Residence	City	5.6	5.5	5.8	94,726	59.4	360.7	<0.001
	Town	7.6	7.3	7.8	44,553	28.0		
	Rural	5.9	5.6	6.3	14,452	9.1		
	Missing	2.5	2.1	2.9	5,664	3.5		
Age group	<25	7.3	7.1	7.6	41,485	26.0	310.9	<0.001
	25-34	5.7	5.6	5.9	78,406	49.2		
	35-44	6.4	6.1	6.7	20,862	13.1		
	45+	4.9	4.6	5.3	13,164	8.3		
	Missing	2.6	2.2	3.0	5,478	3.4		
Sex	Male	10.9	10.4	11.4	14,004	8.8	633.7	<0.001
	Female	5.7	5.5	5.8	143,521	90.0		
	Missing	1.1	0.7	1.6	1,870	1.2		
Marital status	Currently married	6.3	6.1	6.4	60,495	38.0	154.1	<0.001
	Previously married	5.4	5.0	5.9	8,960	5.6		
	Never married	6.2	6.1	6.4	84,613	53.1		
	Missing	2.6	2.2	3.1	5,327	3.3		
Education	Less educated	6.2	6.0	6.4	60,722	38.1	102.0	<0.001
	More educated	6.1	6.0	6.3	93,994	59.0		
	Missing	2.9	2.5	3.5	4,679	2.9		
DHS experience	No	6.1	6.0	6.2	126,437	79.3	154.6	<0.001
	Yes	6.7	6.4	6.9	27,342	17.1		
	Missing	2.7	2.3	3.2	5,616	3.5		
Other experience	No	6.3	6.1	6.5	60,826	38.2	182.2	<0.001
	Yes	6.1	6.0	6.3	92,723	58.2		
	Missing	2.4	2.0	2.8	5,846	3.7		
Permanent employee	No	7.6	7.3	7.8	41,589	26.1	368.3	<0.001
	Yes	7.0	6.7	7.3	30,483	19.1		
	Missing	5.0	4.9	5.2	87,323	54.8		
All	All	6.1	5.9	6.2	159,395	100.0		

Appendix Table A3.4 **Estimates and significance tests for the relationship of the interviewer characteristics to the prevalence of inconsistent ages of women in the household survey and the survey of women.** All surveys are pooled.

Covariate	Category	Prevalence	Lower bound	Upper bound	N	Percent	LRchi2	p-value
Residence	City	3.8	3.7	3.9	94,726	59.4	22.7	<0.001
	Town	4.0	3.8	4.2	44,553	28.0		
	Rural	3.7	3.4	4.0	14,452	9.1		
	Missing	2.8	2.4	3.3	5,664	3.5		
Age group	<25	4.7	4.5	4.9	41,485	26.0	200.7	<0.001
	25-34	3.8	3.6	3.9	78,406	49.2		
	35-44	3.1	2.9	3.4	20,862	13.1		
	45+	2.5	2.2	2.8	13,164	8.3		
	Missing	2.8	2.4	3.3	5,478	3.4		
Sex	Male	3.5	3.2	3.8	14,004	8.8	32.3	<0.001
	Female	3.8	3.8	4.0	143,521	90.0		
	Missing	1.7	1.2	2.4	1,870	1.2		
Marital status	Currently married	3.3	3.2	3.5	60,495	38.0	158.8	<0.001
	Previously married	2.5	2.2	2.9	8,960	5.6		
	Never married	4.3	4.2	4.5	84,613	53.1		
	Missing	2.8	2.4	3.3	5,327	3.3		
Education	Less educated	4.0	3.8	4.1	60,722	38.1	12.9	0.002
	More educated	3.7	3.6	3.8	93,994	59.0		
	Missing	3.1	2.6	3.6	4,679	2.9		
DHS experience	No	3.9	3.8	4.0	126,437	79.3	29.5	<0.001
	Yes	3.4	3.2	3.6	27,342	17.1		
	Missing	2.9	2.5	3.4	5,616	3.5		
Other experience	No	4.2	4.0	4.3	60,826	38.2	47.0	<0.001
	Yes	3.6	3.5	3.8	92,723	58.2		
	Missing	2.7	2.3	3.2	5,846	3.7		
Permanent employee	No	5.4	5.2	5.6	41,589	26.1	372.2	<0.001
	Yes	2.9	2.7	3.1	30,483	19.1		
	Missing	3.4	3.2	3.5	87,323	54.8		
All	All	3.8	3.7	3.9	159,395	100.0		

Appendix Table A3.5 Estimates and significance tests for the relationship of the interviewer characteristics to the prevalence of age heaping in the household survey. All surveys are pooled.

Covariate	Category	Prevalence	Lower bound	Upper bound	N	Percent	LRchi2	p-value
Residence	City	29.3	29.1	29.5	191,464	55.7	130.9	<0.001
	Town	29.4	29.1	29.6	106,795	31.1		
	Rural	27.7	27.2	28.2	32,234	9.4		
	Missing	33.1	32.3	34.0	13,005	3.8		
Age group	<25	27.4	27.1	27.7	86,101	25.1	332.7	<0.001
	25-34	29.7	29.4	29.9	159,971	46.6		
	35-44	31.4	30.9	31.8	50,321	14.6		
	45+	28.5	28.0	29.0	34,443	10.0		
	Missing	32.4	31.6	33.2	12,662	3.7		
Sex	Male	27.4	27.1	27.6	120,105	35.0	662.0	<0.001
	Female	30.1	29.9	30.3	219,228	63.8		
	Missing	43.6	42.1	45.1	4,165	1.2		
Marital status	Currently married	29.8	29.5	30.0	132,425	38.5	119.6	<0.001
	Previously married	29.1	28.5	29.7	19,644	5.7		
	Never married	28.8	28.6	29.0	179,797	52.3		
	Missing	33.2	32.4	34.1	11,632	3.4		
Education	Less educated	29.4	29.1	29.6	126,684	36.9	158.5	<0.001
	More educated	29.0	28.8	29.2	207,160	60.3		
	Missing	35.1	34.1	36.0	9,654	2.8		
DHS experience	No	28.6	28.4	28.8	266,817	77.7	325.2	<0.001
	Yes	31.7	31.4	32.1	64,647	18.8		
	Missing	33.0	32.2	33.9	12,034	3.5		
Other experience	No	28.4	28.1	28.6	122,697	35.7	120.0	<0.001
	Yes	29.7	29.5	29.9	208,101	60.6		
	Missing	32.3	31.5	33.2	12,700	3.7		
Permanent employee	No	31.2	30.9	31.5	99,351	28.9	808.2	<0.001
	Yes	32.0	31.6	32.3	70,775	20.6		
	Missing	27.1	26.9	27.4	173,372	50.5		
All	All	29.3	29.2	29.5	343,498	100.0		

Appendix Table A3.6 Estimates and significance tests for the relationship of the interviewer characteristics to the prevalence of downward displacement below age 15 in the survey of women. All surveys are pooled.

Covariate	Category	Prevalence	Lower bound	Upper bound	N	Percent	LRchi2	p-value
Residence	City	26.9	26.3	27.5	19,273	56.6	11.3	0.010
	Town	25.5	24.6	26.3	10,054	29.5		
	Rural	25.3	23.8	26.8	3,378	9.9		
	Missing	28.1	25.8	30.6	1,372	4.0		
Age group	<25	28.1	27.1	29.0	8,591	25.2	25.2	<0.001
	25-34	25.9	25.2	26.6	16,532	48.5		
	35-44	24.8	23.7	26.0	5,211	15.3		
	45+	25.9	24.2	27.7	2,411	7.1		
	Missing	28.7	26.3	31.2	1,332	3.9		
Sex	Male	26.1	25.4	26.9	11,898	34.9	0.5	0.766
	Female	26.5	25.9	27.1	21,737	63.8		
	Missing	26.2	22.4	30.5	442	1.3		
Marital status	Currently married	24.7	24.0	25.5	12,774	37.5	35.2	<0.001
	Previously married	25.3	23.4	27.2	1,993	5.8		
	Never married	27.6	26.9	28.2	18,092	53.1		
	Missing	28.2	25.8	30.8	1,218	3.6		
Education	Less educated	27.2	26.4	28.0	12,933	38.0	7.5	0.024
	More educated	25.8	25.2	26.5	20,165	59.2		
	Missing	26.8	24.1	29.6	979	2.9		
DHS experience	No	27.0	26.4	27.5	26,627	78.1	33.2	<0.001
	Yes	23.5	22.5	24.6	6,204	18.2		
	Missing	28.2	25.7	30.7	1,246	3.7		
Other experience	No	28.4	27.5	29.2	12,104	35.5	41.8	<0.001
	Yes	25.1	24.5	25.7	20,656	60.6		
	Missing	27.9	25.5	30.4	1,317	3.9		
Permanent employee	No	25.6	24.8	26.5	10,171	29.9	10.6	0.005
	Yes	25.5	24.5	26.6	6,610	19.4		
	Missing	27.1	26.5	27.8	17,296	50.8		
All	All	26.4	25.9	26.9	34,077	100.0		

Appendix Table A3.7 Estimates and significance tests for the relationship of the interviewer characteristics to the prevalence of over-dispersion of ages for children under 5, inferred from the flagging of HAZ and WAZ scores. All surveys are pooled.

Covariate	Category	Prevalence	Lower bound	Upper bound	N	Percent	LRchi2	p-value
Residence	City	14.3	13.9	14.7	29,789	52.2	392.7	<0.001
	Town	9.4	9.0	9.8	19,109	33.5		
	Rural	17.3	16.4	18.2	6,480	11.4		
	Missing	15.6	14.0	17.4	1,689	3.0		
Age group	<25	10.3	9.8	10.9	11,267	19.7	138.7	<0.001
	25-34	14.2	13.8	14.6	28,751	50.4		
	35-44	13.0	12.4	13.7	9,353	16.4		
	45+	11.4	10.6	12.3	5,626	9.9		
	Missing	15.9	14.3	17.5	2,070	3.6		
Sex	Male	15.2	14.7	15.7	18,184	31.9	108.5	<0.001
	Female	12.0	11.7	12.3	38,715	67.8		
	Missing	14.3	9.8	20.4	168	0.3		
Marital status	Currently married	14.0	13.5	14.4	22,192	38.9	54.8	<0.001
	Previously married	13.6	12.7	14.7	4,666	8.2		
	Never married	12.0	11.6	12.4	28,523	50.0		
	Missing	15.7	14.0	17.5	1,686	3.0		
Education	Less educated	12.5	12.1	12.9	23,085	40.5	34.5	<0.001
	More educated	13.5	13.2	13.9	32,803	57.5		
	Missing	8.7	7.3	10.5	1,179	2.1		
DHS experience	No	13.1	12.8	13.4	47,065	82.5	18.5	<0.001
	Yes	12.0	11.3	12.7	8,316	14.6		
	Missing	15.7	14.0	17.5	1,686	3.0		
Other experience	No	10.7	10.4	11.1	26,916	47.2	237.8	<0.001
	Yes	15.0	14.6	15.4	28,449	49.8		
	Missing	15.5	13.9	17.3	1,702	3.0		
Permanent employee	No	11.7	11.2	12.3	13,050	22.9	1,203.9	<0.001
	Yes	6.5	6.1	6.8	16,267	28.5		
	Missing	17.5	17.0	17.9	27,750	48.6		
All	All	13.0	12.7	13.3	57,067	100.0		