USAID
FROM THE AMERICAN PEOPLE

MODELING INTERVIEWER EFFECTS IN DHS SURVEYS

DHS METHODOLOGICAL REPORTS 32

AUGUST 2022

This publication was produced for review by the United States Agency for International Development. It was prepared by Mahmoud Elkasabi and Azam Khan.

DHS Methodological Reports No. 32

Modeling Interviewer Effects in DHS Surveys

Mahmoud Elkasabi[1,2]
Azam Khan[1]

ICF
Rockville, Maryland, USA

August 2022

[1] ICF
[2] The DHS Program

Corresponding author: Mahmoud Elkasabi, The Demographic and Health Surveys Program, ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA; phone: +1 301-572-0484; fax: +1 301-572-0999; email: mahmoudelkasabi@gmail.com

Acknowledgments: The authors thank Joanna Lowell, Anjushree Pradhan, Yodit Bekele, Suzanne Arrington, Sara Scates, and Natalia Woolley for helping with classifying the study questions. The authors also thank Thomas Pullum for his feedback on the data analysis presented in this report. Finally, the authors thank Brady West for very helpful comments on an earlier draft of this report.

Editor: Diane Stoy
Document Production: Joan Wardell, Natalie Shattuck, and Chris Gramer

This study was conducted with support from the United States Agency for International Development (USAID) through The DHS Program (#720-OAA-18C-00083). 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: +1 301-407-6500, fax: +1 301-407-6501, email: info@DHSprogram.com, internet: www.DHSprogram.com.

Recommended citation:

Elkasabi, Mahmoud and Azam Khan. 2022. *Modeling Interviewer Effects in DHS Surveys*. DHS Methodological Reports No. 32. Rockville, Maryland, USA: ICF.

CONTENTS

TABLES

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 our hope 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

A previous DHS methodological report (MR24) examined the effects of interviewer characteristics on data quality in DHS surveys. That report examined if variation in 25 indicators of data quality, across 15 DHS surveys, could be attributed to the interviewers and their characteristics. According to MR24, interviewers who are older and better educated have lower levels of problematic outcomes, while interviewers with prior experience with a DHS survey or other surveys are often associated with statistically significant outcomes that are often in favor of better quality data. The results of MR24 did not account for the interviewer assignments to sampling clusters, where interviews are typically nested within a cross-classification of sampling clusters and interviewers. Moreover, the results did not control for the respondent characteristics.

As an extension of that effort, the current report uses multilevel models to estimate interviewer effects in DHS surveys, while accounting for the structure of the interviewer assignments and the characteristics of both respondents and interviewers. Based on data from 24 recent DHS surveys and more than 100 questions from the Woman's Questionnaire in each survey, this report examines interviewer effects across countries and across different characteristics of questions, such as length (longer versus shorter questions), sensitivity (questions on sensitive topics versus questions on non-sensitive topics), social desirability (questions prone to social desirability bias versus questions not prone to social desirability bias), complexity and/or difficulty (complex or difficult questions versus questions that are not complex or difficult), and question type (whether the information collected by the question was factual or non-factual). Long questions, non-factual questions, and questions on complex or difficult topics were associated with larger interviewer effects compared to the shorter questions, factual questions, and questions on less complex or difficult topics. These differences were consistent across most surveys.

The analysis in this report can be extended to additional questions and surveys in the future. Results from these analyses can improve the quality of interviews and data collected by improving training for interviewers before fieldwork and monitoring interviewer performance during fieldwork.

Key words: data quality, interviewer effects, question characteristics.

ACRONYMS AND ABBREVIATIONS

CAPI	computer-assisted personal interviewing
DHS	Demographic and Health Surveys
FW	Fieldworker Questionnaire
ICC	intraclass correlation coefficient
IIC	intra-interviewer correlation
IQR	interquartile range
IR	Individual Recode
NHIS	National Health Interview Survey
STI	sexually transmitted infection

COUNTRY CODES USED IN THIS REPORT

AM	Armenia
BD	Bangladesh
BJ	Benin
BU	Burundi
CM	Cameroon
GM	The Gambia
GN	Guinea
HT	Haiti
LB	Lebanon
ML	Mali
MM	Myanmar
MW	Malawi
NG	Nigeria
NP	Nepal
PH	Philippines
PK	Pakistan
RW	Rwanda
SL	Sierra Leone
TJ	Tajikistan
TL	Timor-Leste
UG	Uganda
ZA	South Africa
ZM	Zambia
ZW	Zimbabwe

1 BACKGROUND

The Demographic and Health Surveys (DHS) are key sources of population and health indicators in many countries. The surveys are designed to collect representative data that enable high quality population estimates of key indicators related to fertility, family planning methods, maternal and child health, as well as childhood and adult mortality, HIV/AIDS and other sexually transmitted infections (STIs), women's empowerment, and domestic violence. To maintain and improve the quality of the DHS data, it is important to investigate the factors that affect data quality and improve survey processes, systems, and methodologies.

In a typical DHS data collection, several questionnaires are used, such as the Household Questionnaire, the Woman's Questionnaire, and the Man's Questionnaire. The Woman's Questionnaire is the key tool used to collect data for most key DHS indicators. Therefore, the data collected with this questionnaire are the focus of this report.

Conducting surveys through personal interviews involves a complex interaction between the respondent and the interviewer. This interaction is often affected by the survey tools, such as the questionnaire, and the environment or context in which the interview is conducted. Interviewers have been shown to have a significant impact on the quality of data generated from a survey (Olson et al. 2020). This can either have a positive or negative effect on the data. In the DHS, interviewers are organized into teams that are assigned to work in more than one sampled cluster. Since an interviewer could potentially interview respondents across clusters (conducting interviews in more than one cluster), responses are nested within a cross-classification of clusters and interviewers.

Interviewer effects can have many potential causes, one of which relates to the characteristics of the questions administered by the interviewer. Questions that may be viewed as sensitive, or questions on embarrassing topics, have been shown to be more prone to interviewer effects (Mangione, Fowler, and Louis 1992; Schnell and Kreuter 2005). Difficult questions or topics of a complex nature may also be prone to interviewer effects because these questions may be more complicated to administer and may elicit requests for clarification and probing. The type of information collected from the question could also be prone to interviewer effects. Subjective questions such as those related to perceptions and attitudes have larger interviewer effects compared to factual questions such as respondent demographics. Finally, lengthy questions are also associated with larger interviewer effects (Mangione, Fowler, and Louis 1992; Pickery and Loosveldt 2001; Schnell and Kreuter 2005; West and Blom 2017).

The DHS Program published a methodological report (MR24) that studied the interviewer effects in the DHS surveys within an assessment of data quality (Pullum et al. 2018). In MR24, the authors looked at the effects of interviewer characteristics on data quality in DHS surveys by examining if variation in 25 indicators of data quality across three broad categories (non-response and refusals, reported age at death of young children, and ages and dates) could be attributed to interviewer characteristics. The study found that for many indicators, a large portion of the variance could be attributed to a handful of interviewers. In addition, there are a number of published papers that have examined interviewer effects in DHS data, although most have focused on specific DHS topics or modules and/or specific countries, such as the study of interviewer effects in sensitive questions within the domestic violence module in the India DHS surveys (Singh, Kumar, and Arnold 2022), the study of interviewer effects on abortion data (Footman 2021; Leone, Sochas, and Coast 2021), and the study of interviewer effects on data of contraceptive use in Indonesia and Philippines (Amos 2018).

In this report, we use cross-classified multilevel models to analyze interviewer variance and estimate the impact of interviewer effects in the DHS Woman's Questionnaire. We first use multilevel models adjusting for the characteristics of respondents (age, marital status, and education level) and sampling clusters (residence type, and geographic region) that define the interviewer assignments to account for the amount of clustering in survey responses by interviewer. We then account for interviewer characteristics (age, marital status, education level, and previous experience with DHS surveys and other surveys) in the multilevel model, before finally examining how interviewer effects vary by question characteristics.

This study provides the first assessment of the increase in estimated variance due to interviewer effects and its impact on the data collected in the DHS Woman's Questionnaire. In this report, we included as many questions and countries as possible. The report includes data from 24 recent DHS data collections and more than 100 questions in each survey. We included most DHS studies with accessible data about interviewer characteristics, and selected questions that represent most sections and question types in the DHS Woman's Questionnaire. We identified the characteristics of all questions and explored the relationships between question characteristics and interviewer effects. Chapter 2 of the report describes the data and methods. Chapter 3 presents the results, and Chapter 4 concludes with a discussion.

2 DATA AND METHODS

2.1 Data

We used data from 24 recent DHS surveys conducted between 2015 and 2020, as listed in Table 1. In these surveys, we used data from the Woman's Questionnaire, available in the Individual Recode (IR) datasets. We also use interviewer-level data available in the Fieldworker Questionnaire (FW) datasets. Both the IR and FW datasets are available on the DHS website (https://dhsprogram.com/). In these 24 surveys, interviews with 6,116 to 41,821 women, age 15-49, were collected from a range of 280 to 1,389 sampling clusters (census enumeration areas). The interviews were collected by 52 to 269 female interviewers, depending on the survey. As in all DHS surveys, interviewers in these surveys worked in teams of three to six interviewers. Each team is responsible for visiting selected clusters and completing the DHS questionnaires with pre-selected households.

As shown in Figure 1, any given cluster is assigned to only one team, and each team is responsible for more than one cluster. In most cases, teams are assigned to work in clusters in a specific geographic region. Figure 2 shows that the median number of interviews collected by interviewers ranged from 100 to 300 interviews, except in Haiti (HT) and Nigeria (NG), where the median exceeded 300 interviews per interviewer, and in South Africa (ZA) where the median was 92. With the dispersion of the number of interviews per interviewer, the interquartile range (IQR)—the difference between the third quartile and the first quartile—indicated that the data are very dispersed only in few countries, such as Myanmar (MM), Nigeria (NG), Pakistan (PK), Sierra Leone (SL) and Zimbabwe (ZW), where the IQR values were 130, 157, 174, 268, and 198, respectively.

Across most countries, a median of 15 to 30 clusters per interviewer were visited, except in Bangladesh (BD), Nigeria (NG), and Uganda (UG), where the medians were 35, 36 and 32, respectively, and in Philippines (PH), where the median was 12 (see Figure 3). With the dispersion of the number of clusters per interviewer, the IQR indicates that data are more dispersed in Armenia (AM), Bangladesh (BD), Myanmar (MM), Pakistan (PK), and Zimbabwe (ZW), where the IQRs were 12, 13, 11, 15, and 20, respectively, compared to other countries where the IQR was 8 or less.

With the cluster-level workload, or the average number of interviews per cluster completed by any given interviewer, the median number of interviews per cluster ranged from 7 and 11 interviews in most surveys, except in Gambia (GM) and Sierra Leone (SL), where the medians were 13 and 12, respectively, and in Armenia (AM), Bangladesh (BD), and South Africa (ZA), where the medians were 5, 6, and 4, respectively (see Figure 4). For the dispersion of the workloads, the IQR indicates that workloads were relatively more dispersed in Sierra Leone (SL) and Zimbabwe (ZW), where the IQR were 10 and 6, respectively, compared to other countries where the IQR was 4 or less.

Table 1 DHS surveys in this study

Country/year	Clusters	Interviews	Interviewers
Armenia 2015–16	313	6,116	57
Bangladesh 2017–18	672	20,127	109
Benin 2017–18	555	15,928	73
Burundi 2016–17	554	17,269	77
Cameroon 2018	429	14,677	68
Gambia 2019–20	280	11,865	52
Guinea 2018	401	10,874	61
Haiti 2016–17	450	15,513	55
Liberia 2019–20	325	8,065	55
Malawi 2015–16	850	24,562	148
Mali 2018	345	10,519	71
Myanmar 2015–16	441	12,885	84
Nepal 2016	383	12,862	57
Nigeria 2018	1,389	41,821	118
Pakistan 2017–18	561	15,068	92
Philippines 2017	1,248	25,074	269
Rwanda 2019–20	500	14,634	59
Sierra Leone 2019	576	15,574	68
South Africa 2016	729	8,514	91
Tajikistan 2017	366	10,718	64
Timor-Leste 2016	455	12,607	80
Uganda 2016	696	18,506	91
Zambia 2018	545	13,683	70
Zimbabwe 2015	400	9,955	69

Figure 1 Assignment of teams to sample clusters in DHS surveys

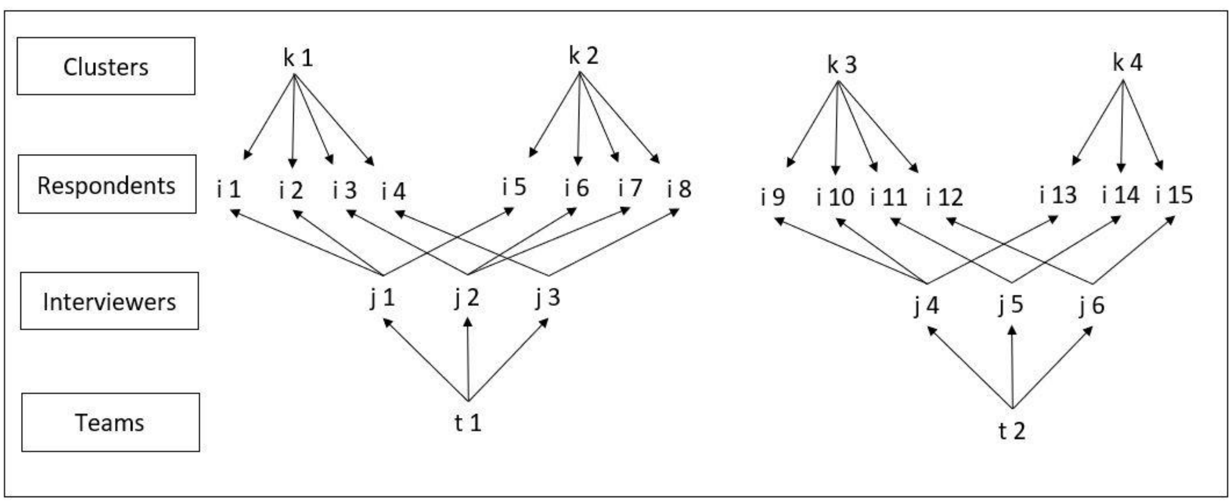
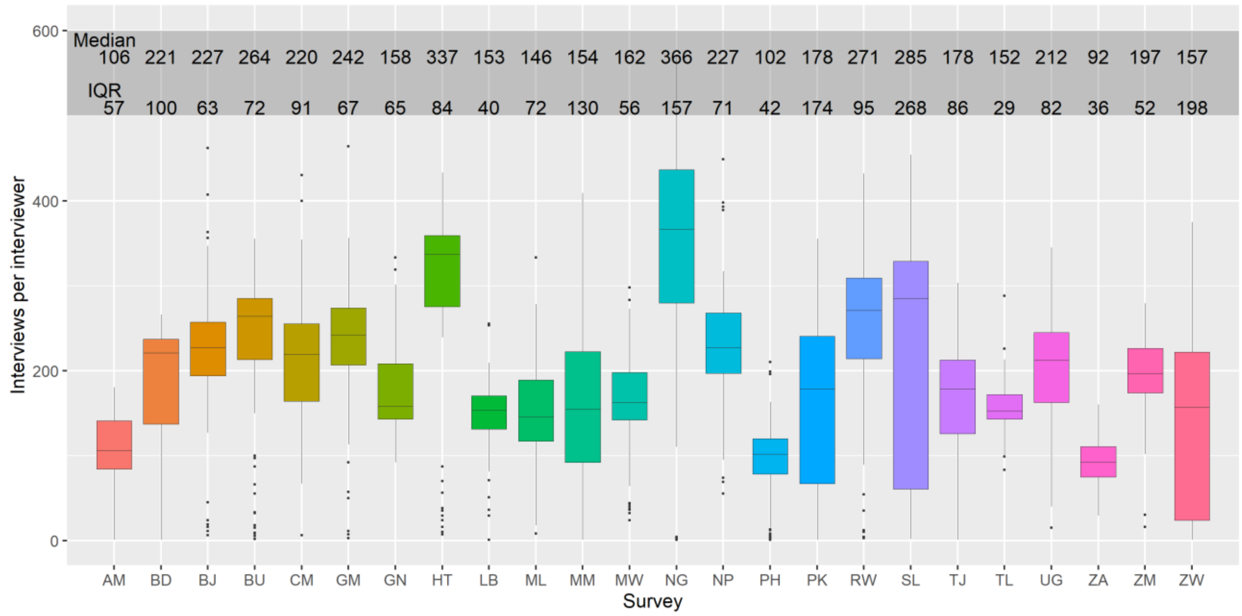
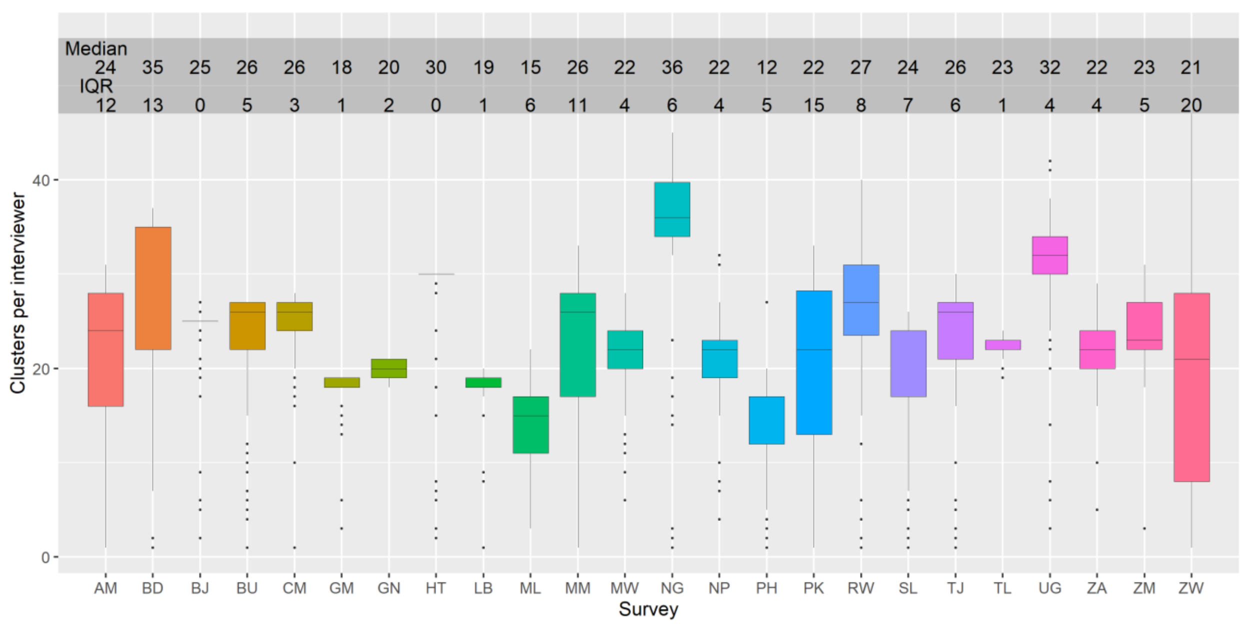


Figure 2 Number of interviews per interviewer, by survey



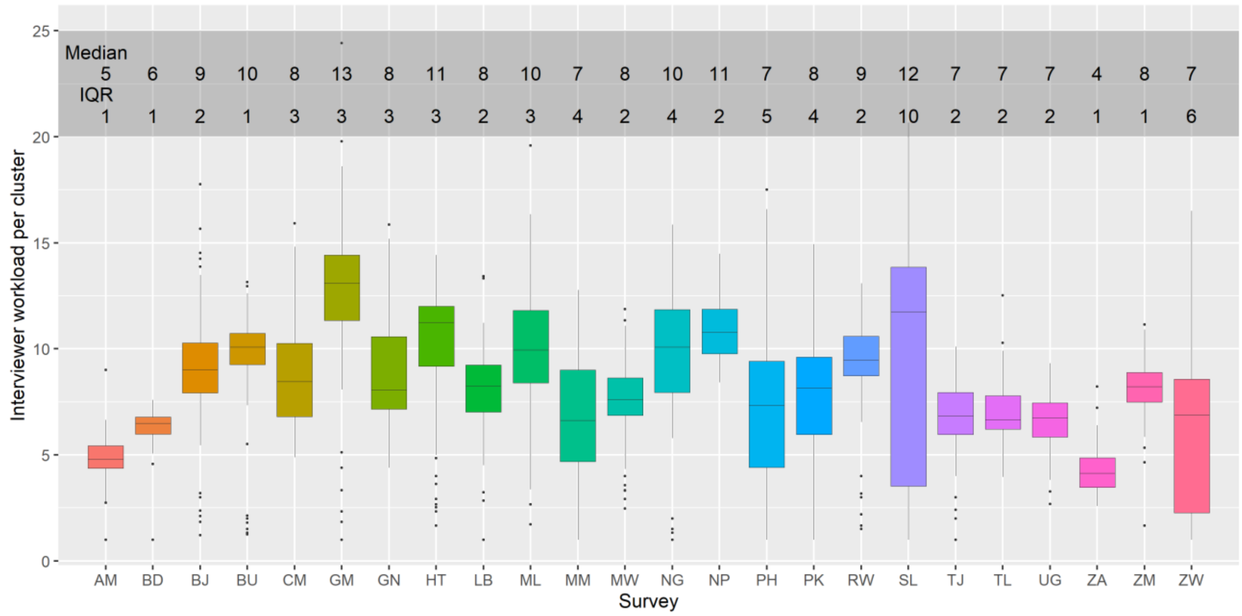
Note: See page xv for a list of country codes used in this report.

Figure 3 Number of clusters worked per interviewer, by survey



Note: See page xv for a list of country codes used in this report.

Figure 4 Interviewer workload per cluster, by survey



Note: See page xv for a list of country codes used in this report.

2.2 Methods

2.2.1 Questions, variables, and outcomes

This study considered a subset of questions from the DHS VII Woman's Questionnaire. To facilitate a more comprehensive assessment, the selected questions included a broad range of topics for a total of 116 questions across nine of the eleven sections of the DHS VII Woman's Questionnaire (Table 3). The selected questions map to about 102 dichotomous outcomes per country, which form the basis of this research. These dichotomous outcomes are based on dichotomous, nominal, or ordinal variables from the DHS woman's datasets (IR). One dichotomous variable makes one outcome variable, while a nominal or ordinal variable makes multiple dichotomous outcomes. For example, "Do you read a newspaper or magazine at least once a week, less than once a week, or not at all?" is coded in v157 that is an ordinal variable with 3 response categories (0: *not at all*; 1: *less than once*; 2: *at least once*). This variable creates 3 dichotomous outcomes as follows: 1) v157_0: *does not read newspapers or magazines at all (Yes/No)*; 2) v157_2: *reads newspapers or magazines less than once a week (Yes/No)*; 1) v157_3: *reads newspapers or magazines at least once a week (Yes/No)*. Up to 50 outcomes are based on 50 dichotomous variables, and the remaining 66 nominal or ordinal variables create the remaining dichotomous outcomes in each country. To avoid outcomes with rare prevalence, only response categories with a prevalence of 5% or greater were made into separate outcomes. Response categories with prevalence less than 5% were combined with other response categories. Dichotomous variables with prevalence less than 5%, or greater than 95%, were not considered as outcomes in this analysis.

The same approach was used in other studies such as Mangione, Fowler, and Louis (1992) and Dahlhamer et al. (2020). The standardized variable names and definitions in the IR data files facilitated comparisons across DHS surveys. Table 2 presents the number of outcomes by survey country. Table 3 shows the number of questions by questionnaire section included in this analysis.

We did not include any questions from Sections 5 and 6, which cover child immunization, child health, and child nutrition. In these sections, detailed data about vaccinations and nutrition are collected about children under age 5. This involves collecting data that are recorded on vaccination cards or based on the mother's memory. We believe that these sections are worthy of a separate study to examine the interviewer effects on the questions. We also considered the core Woman's Questionnaire only; optional modules such as domestic violence were not considered. Table A.1 in the appendix provides the full list of questions and potential outcomes (outcomes before combining the rare outcomes) included in this analysis.

Table 2 Outcomes by DHS surveys in this study

Country/year	Outcomes
Armenia 2015–16	100
Bangladesh 2017–18	57
Benin 2017–18	109
Burundi 2016–17	105
Cameroon 2018	108
Gambia 2019–20	106
Guinea 2018	105
Haiti 2016–17	107
Liberia 2019–20	106
Malawi 2015–16	110
Mali 2018	106
Myanmar 2015–16	90
Nepal 2016	103
Nigeria 2018	102
Pakistan 2017–18	99
Philippines 2017	101
Rwanda 2019–20	105
Sierra Leone 2019	106
South Africa 2016	89
Tajikistan 2017	104
Timor-Leste 2016	96
Uganda 2016	108
Zambia 2018	106
Zimbabwe 2015	109

Table 3 Distribution of selected questions by questionnaire section

Questionnaire section	Questions
Section 1: Respondent's Background	8
Section 2: Reproduction	11
Section 3: Contraception	18
Section 4: Pregnancy and Postnatal Care	14
Section 7: Marriage and Sexual Activity	4
Section 8: Fertility Preferences	9
Section 9: Husband's Background and Woman's Work	17
Section 10: HIV/AIDS	28
Section 11: Other Health Issues	7
Total	**116**

2.2.2 Question characteristics

The specific questions included in the analysis were selected purposively based on question characteristics that were associated with interviewer effects from past research.[1] There were five question characteristics considered for this analysis:

1) **Length of the question (long versus short questions):** Longer questions may lead to larger interviewer effects compared to shorter questions (Dahlhamer et al. 2020; Mangione, Fowler, and Louis 1992; Pickery and Loosveldt 2001). The hypothesis is that longer questions may result in more opportunities for the interviewer to improvise and deviate from the standard interviewing protocol when asking the question. In addition, respondents may experience greater comprehension problems with longer questions because more information must be processed in order to respond to the question. This may elicit more requests for clarification or repeating the question, which result in additional opportunities for improvisation and deviation from the standard interviewing protocol by the interviewer.

2) **Type of question (factual versus non-factual, or attitudinal/subjective):** Questions that are subjective, such as attitudinal or opinion-related questions, may be more prone to interviewer effects compared to objective questions such as factual and demographic questions (West and Blom 2017). In responding to subjective questions, respondents might be influenced by interviewer characteristics and/or the environment to respond in a certain way compared to factual questions.

3) **Question sensitivity:** Sensitive questions may lead to larger interviewer effects compared to non-sensitive questions (Mangione, Fowler, and Louis 1992; Schnell and Kreuter 2005). When responding to sensitive questions, respondents may be more prone to interviewer effects than when responding to non-sensitive or factual questions.

4) **Social Desirability:** This domain considers questions that may elicit social desirability bias in respondents. Social desirability bias is the tendency of respondents to answer questions in a manner that may be viewed favorably by the interviewer. This may lead to an under-reporting of socially undesirable attitudes and/or behaviors and over-reporting of socially desirable attitudes and/or behaviors (Davis et al. 2009). For questions that are more likely for respondents to feel the need to give a socially desirable response, interviewer effects may be greater, as compared to questions that are less likely for respondents to feel the need to respond in a socially desirable way.

5) **Difficulty/Complexity:** Difficult questions or questions on complex topics are questions that require respondents a) to answer on complicated topics that they may have given little attention or thought or b) to recall events and/or behaviors that may be difficult to remember. Those questions may be more prone to interviewer effects (Dahlhamer et al. 2020; Mangione, Fowler, and Louis 1992; Pickery and Loosveldt 2001). We hypothesize that these questions may generate a greater number of inadequate responses and/or more requests for clarification by the respondent, which

[1] This study does not attempt to provide an exhaustive or representative assessment of interviewer effects for the DHS VII Woman's Questionnaire, but highlights the types of question characteristics that could potentially be associated with greater interviewer effects in this questionnaire.

may in turn provide more opportunities for the interviewers to improvise and deviate from the interviewing protocol, which would potentially lead to larger interviewer effects.

2.2.3 Question classification by characteristics

To classify questions according to the "length of question," we used the number of characters in the question as a proxy for length, not including spaces and punctuations.[2] The number of characters in each question was computed. Using the distribution of the count of characters across all questions, we then determined four discrete categories using quartiles as cut points. Questions that fell in the third and fourth quartile were considered long questions (with 83 or more characters in length).[3]

The classification of questions as factual or non-factual considered the type of information collected by the question. If the information was of a subjective nature, such as attitudinal or opinion-related responses, the question was non-factual. Other questions that collect information of an objective or non-subjective nature, such as demographics, were considered factual.

To identify questions according to sensitivity, social desirability, and difficulty/complexity, we followed an approach similar to that described by Dahlhamer et al. (2019, 2020). Six DHS survey managers (three of whom with more than 10 years of experience with DHS surveys) rated each study question based on a series of rating items for each characteristic.

For the sensitivity dimension, the six DHS managers were asked to respond to the following three rating items using a 5-point scale (1 = completely disagree to 5 = completely agree):

- This question is very personal.
- I would be uncomfortable asking this question to a respondent.
- I would be uncomfortable responding to this question with an interviewer present.

After examining the reliability of the answers to each item (Table 4), the ratings for each question were then summed across the three items and the six raters to produce an overall index of sensitivity per question. Using the distribution of the index across all questions, four discrete categories with the quartiles as cut points were determined. Questions that fell in quartiles three and four had the highest sensitivity scores——index scores ranging from 35 to 77—and were considered sensitive questions.

A similar approach was used to determine questions that may be more prone to social desirability bias. The six raters were asked to rate each question with a 3-point scale (1 = Not at all likely, 2 = Somewhat likely, 3 = Very likely) using the following rating item:

- How likely is this question to elicit a more favorable response from the respondent to the interviewer?

[2] We acknowledge some limitations with this approach. Other measures such as the number of words could also be used. We also acknowledge that words may not have equal value in a question, where some words may be more important or carry more weight than others. In addition, the character count is based on the English questions and will vary with other languages.

[3] The count of characters includes all words in the question, as well as any additional text, such as probes.

The ratings for each question where then summed across the raters to produce an overall index of social desirability per question. Using the distribution of the index across all questions, four discrete categories using the quartiles as cut points were determined. Questions that fell in quartiles three and four had the highest social desirability scores with index scores that ranged from 9 to 17 and were considered questions that are more prone to social desirability bias.

Finally, to determine difficult/complex questions, the raters rated each question based on the following rating items using a 5-point scale (1 = completely disagree to 5 = completely agree):

- This question requires respondents to recall behaviors or events that may be difficult to remember.
- This question may be complicated for a respondent to understand.
- This question covers a complex topic for which the respondent may usually give little or no thought.

After examining the reliability of the answers to each item (See Table 4), we dropped the second item because the answers were unreliable. Therefore, the ratings for each question were then summed across the two remaining items and the six raters to produce an overall difficulty index. Using the distribution of the index across all questions, four discrete categories using the quartiles as cut points were determined. Questions that fell in quartiles three and four had the highest scores (index scores from 35 to 68) and were considered difficult or complex questions.

To assess the reliability of the answers to the rating items, before creating the indexes of all characteristics, we examined the inter-rater reliability across different rating items for each characteristic. For a measure of inter-rater reliability, we calculated the intraclass correlation coefficient (ICC) (Koo and Li 2016), using the icc function of the irr R package.

Item indexes were calculated based on items with an ICC of 70% or more. Table 4 presents the ICC by items and flags items used in the index calculations.

Table 4 Inter-rater reliability ICC by items

Characteristics	Item	ICC %	Item used for index calculation
Sensitivity	This question is very personal.	84	√
	I would be uncomfortable asking this question to a respondent.	72	√
	I would be uncomfortable answering this question to an interviewer.	75	√
Social desirability	How likely is this question to elicit a more favorable response from the respondent to the interviewer?	81	√
Difficulty/complexity	This question requires respondents to recall behaviors or events that may be difficult to remember.	84	√
	This question may be complicated for respondent to understand.	40	X
	This question covers a complex topic which the respondent may usually give little or no thought to.	73	√

Table 5 provides the frequency and percentage distributions of each question by the characteristics described above.

Table 5 Distribution of questions according to characteristics

Characteristics		Questions	%
Length	**Not long**	57	49.1
	Long	59	50.9
Type	**Factual**	78	67.2
	Non-factual	38	32.8
Sensitivity	**Yes**	63	54.3
	No	53	45.7
Social Desirability	**Yes**	79	68.1
	No	37	31.9
Difficulty/Complexity	**Yes**	58	50.0
	No	58	50.0

Table A.1 in the appendix presents the questions and variables used in the analysis by question characteristics.

2.2.4 Estimating intra-interviewer correlation

Due to the cross-classified structure illustrated in Figure 1, outcomes measured from respondents (level 1) are uniquely nested within cross-classifications of interviewers (level 2) and clusters (level 2). For each of the outcomes, we fit five mixed-effects logistic regression models to model $\pi_{i(j,k)} = Pr\left(y_{i(j,k)} = 1\right)$ for outcome *y* measured by interviewer *j* from respondent *i* in cluster *k*. In the five models, we added the random effects of interviewers as follows:

1) An unconditional model that included only a random effect due to interviewers u_{oj} where $u_{oj} \sim N\left(0, \sigma^2_{u_j}\right)$:

$$ln\left(\pi_{i(j,k)}\right) = \beta_o + u_{oj}$$

2) A model that included random effects due to interviewers u_{oj} and random effects due to clusters u_{ok} where $u_{ok} \sim N\left(0, \sigma^2_{u_k}\right)$:

$$ln\left(\pi_{i(j,k)}\right) = \beta_o + u_{oj} + u_{ok}$$

3) In addition to the random effects in step 2, this model included fixed effects of a set of *A* respondent characteristics X^I_{aijk} measured for each respondent *i* by interviewer *j* in cluster *k*:

$$ln\left(\pi_{i(j,k)}\right) = \beta_o + \sum_{a=1}^{A} \beta_a X^I_{aijk} + u_{oj} + u_{ok}$$

4) In addition to the random and fixed effects in Step 3, this model included fixed effects of a set of *B* cluster characteristics X^K_{bk} measured for each cluster *k*:

$$ln(\pi_{i(j,k)}) = \beta_o + \sum_{a=1}^{A} \beta_a X_{aijk}^{I} + \sum_{b=1}^{B} \beta_b X_{bk}^{K} + u_{oj} + u_{ok}$$

5) Finally, this model added fixed effects of a set of *C* interviewer characteristics X_{cj}^{J} measured for each interviewer *j*:

$$ln(\pi_{i(j,k)}) = \beta_o + \sum_{a=1}^{A} \beta_a X_{aijk}^{I} + \sum_{b=1}^{B} \beta_b X_{bk}^{K} + \sum_{c=1}^{C} \beta_c X_{cj}^{J} + u_{oj} + u_{ok}$$

In our results section, we present results from the final model, and the intra-interviewer correlation (IIC), which expresses the ratio of the between-interviewer variance in an outcome variable to the total variance in the same variable. Since all outcome variables are dichotomous, the following formula approximated the value of the IIC:

$$IIC = \frac{\sigma_{u_j}^2}{\sigma_{u_k}^2 + \sigma_{u_j}^2 + 3.29}$$

In this equation, the respondent variance (level 1) is set at 3.29, which is the variance of the underlying standard logistic distribution (Snijders and Bosker 1999). All models were estimated with the *glmer* function of the *lme4* R package.

2.2.5 Covariates included in multilevel models

Respondent characteristics included age (15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49), marital status (never in union, married, living with partner, widowed, divorced, or separated), education (no education, primary, secondary, higher), and if a translation was used during the interview (Yes or No).

Interviewer characteristics included age (15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65+), marital status (never in union, married, living with partner, widowed, divorced, or separated), education (secondary, higher), experience with previous DHS surveys (Yes or No), and experience with other surveys (Yes or No).

Cluster characteristics included region (geographic regions vary by country) and residence type (urban, rural).

We did not include gender of interviewers and respondents as covariates since all respondents and interviewers are females. We also did not include the survey weights as a covariate in the final model, because it did not lead to any significant changes in the model results when it was added to an earlier version of the model.

2.2.6 Data analysis of intra-interviewer correlation estimates

The IIC estimates were compared across different characteristics (long versus not long, factual vs. not factual, sensitive versus non-sensitive, socially desirable versus not socially desirable, and difficult versus not difficult), and were analyzed on the question level. For any question coded as a nominal or ordinal variable where multiple outcomes were generated, the question was assigned an IIC value that is the average of IICs for the outcomes of that question. This approach was used so that the results would not be dominated by results of questions with multiple outcomes.

We computed the median and IQR of IICs in most comparisons. We made comparisons across question characteristics based on all data from all surveys and across different surveys. In presenting these comparisons, we used boxplots and annotated values of medians and IQR in most graphs. To test for significant differences in median IICs by characteristics, we used the following non-parametric tests: the Mann–Whitney–Wilcoxon two-sample test for characteristics with two categories, and the Kruskal–Wallis test for measures with three or more categories (Dahlhamer et al. 2020). For each comparison, we indicated different significance levels for *p* values. To test for significant differences in IICs after controlling for all question characteristics, we used Beta regression to model IICs with the five characteristics as covariates.[4] We fit a separate model for each survey, and then compared results across surveys.

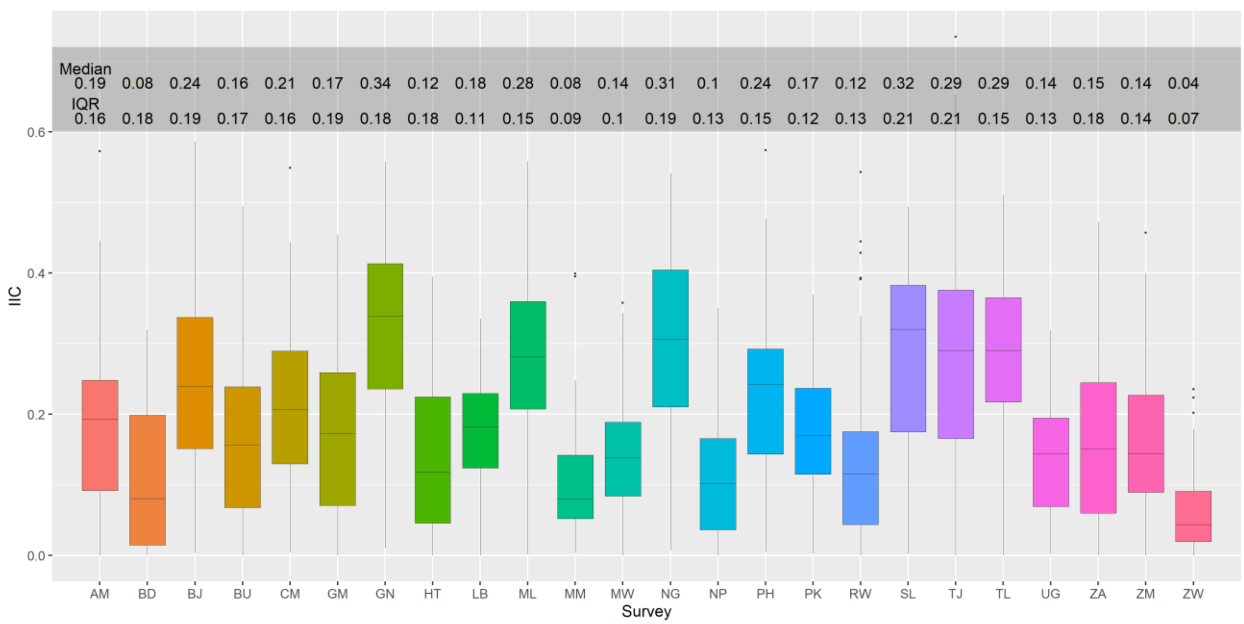
[4] Beta regression was selected because it is suitable for modeling dependent variables that are percentages—IIC in this case.

3 RESULTS

3.1 Intra-interviewer Correlations by Survey

Figure 5 presents boxplots of IICs across countries. Medians and IQRs for the IICs are annotated above the boxplots. The median IIC ranged from 0.04 in Zimbabwe (ZW) to 0.34 in Guinea (GN). The median IICs were less than 0.25 in all countries, except in Guinea (GN), Mali (ML), Nigeria (NG), Sierra Leone (SL), Tajikistan (TJ), and Timor-Leste (TL), where median IICs were 0.34, 0.28, 0.31, 0.32, 0.29, and 0.29, respectively. For almost all countries, the IQR of IICs ranged between 0.10 and 0.20, except for Myanmar (MM) and Zimbabwe (ZW), where the IQRs were 0.09 and 0.07, respectively. These results indicate that interviewer effects tend to vary significantly across countries.

Figure 5 IIC by country

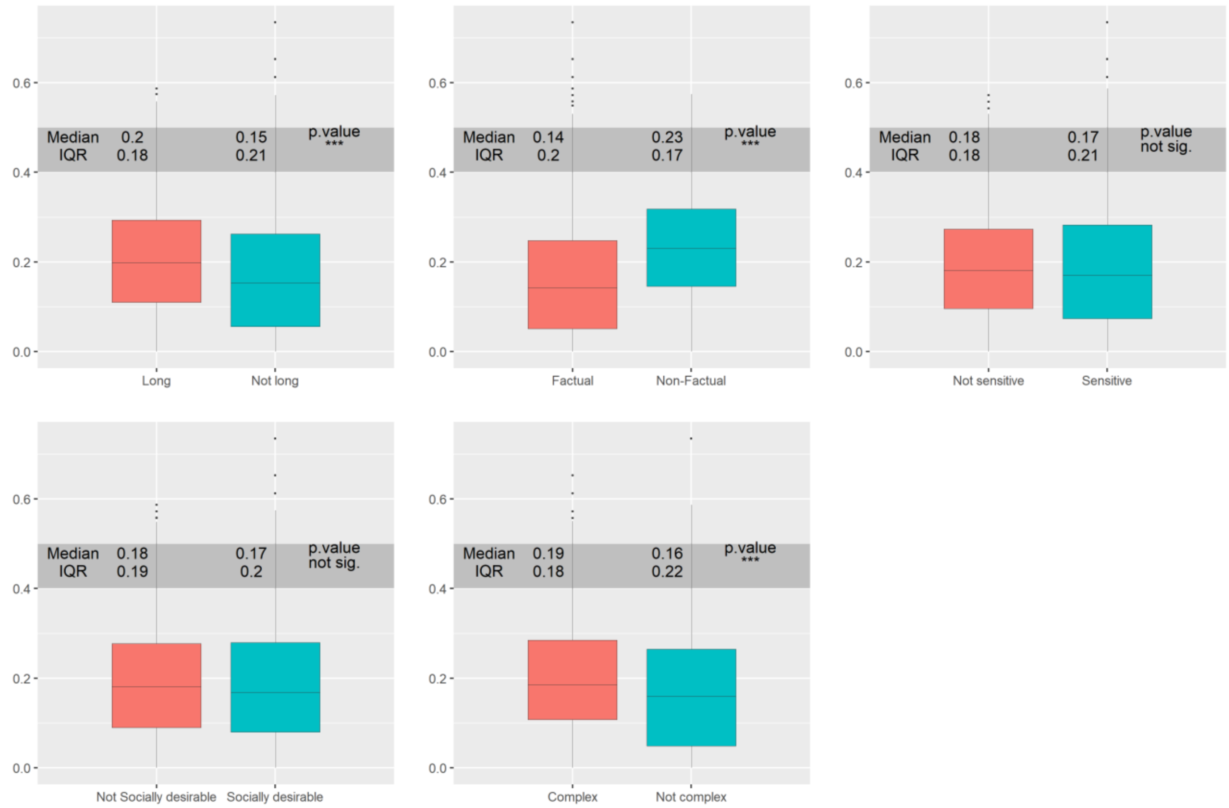


Note: See page xv for a list of country codes used in this report.

3.2 Intra-interviewer Correlations by Question Characteristics

Figure 6 presents boxplots of IICs according to five question characteristics. Three of the characteristics were significantly associated with interviewer effects: length of questions, question type (factual versus non-factual), and question complexity. The median IIC of long questions (0.2) is significantly higher than the median IIC of shorter questions (0.15) (Mann–Whitney–Wilcoxon test, p value ≤ .001). The median IIC of non-factual questions (0.23) is significantly higher than the median IIC of factual questions (0.14) (Mann–Whitney–Wilcoxon test, p value ≤ .001). The median IIC of complex/difficult questions (0.19) is significantly higher than the median IIC of less complex/difficult questions (0.16) (Mann–Whitney–Wilcoxon test, p value ≤ .001).

Figure 6　IIC by question characteristics

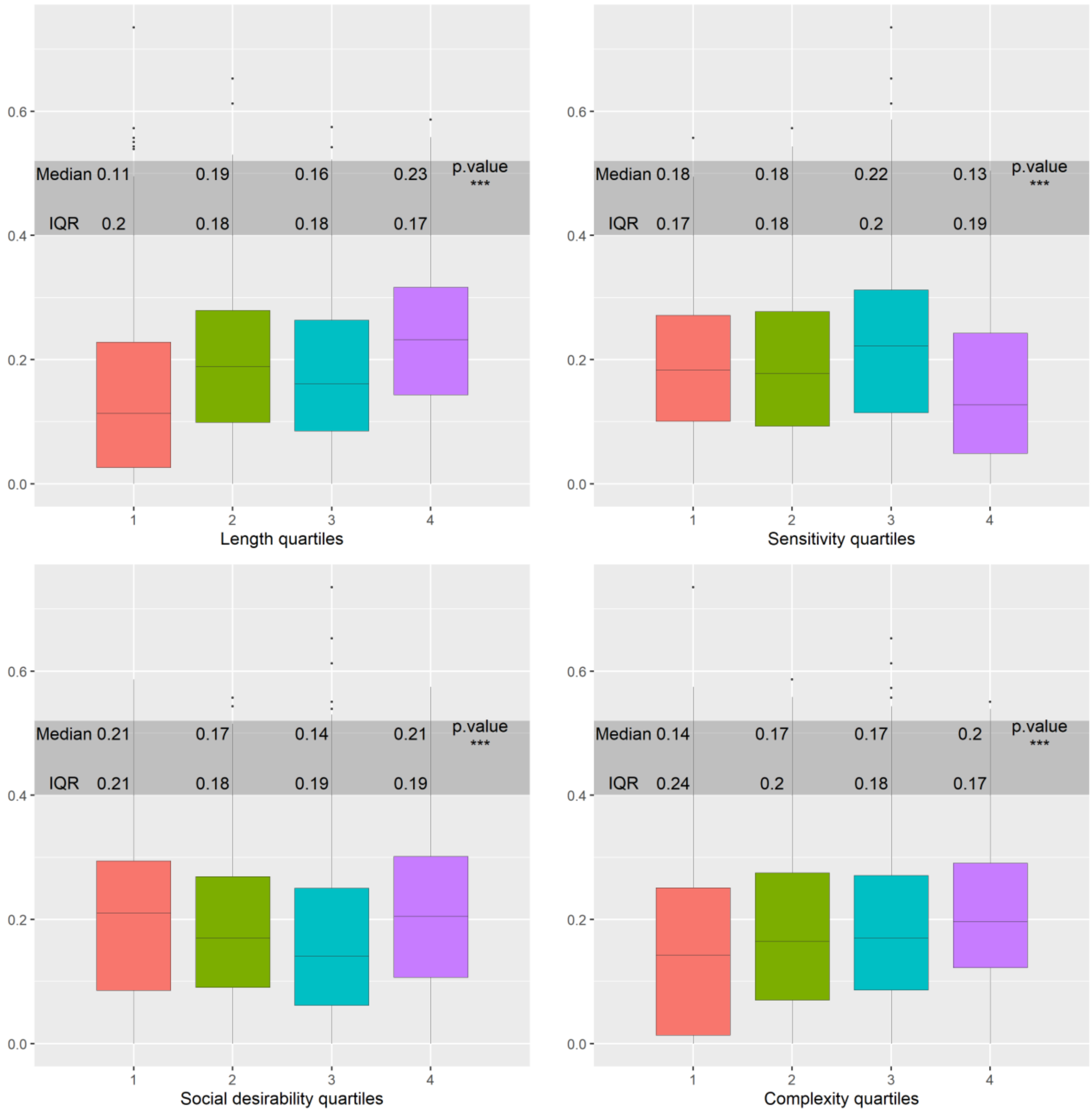


*** Significant at $p \leq .001$; ** significant at $.001 < p \leq .05$; * significant at $.05 < p \leq .1$; and not significant at $p > .1$ for the Mann–Whitney–Wilcoxon test

Figure 7 presents the boxplots of IICs across quartiles of the four question characteristics: length, sensitivity, social desirability, and complexity/difficulty. No quartiles are available for the question type (factual versus non-factual) because it was not assigned based on an index. For the characteristics of length and complexity/difficulty, the general trend shows an increase in median IICs for higher quartiles. The differences are quite apparent between median IICs of the first and last quartiles (length: quartile 1 median IIC = .11; quartile 2 median IIC = .19; quartile 3 median IIC = .16; quartile 4 median IIC = .23; complexity: quartile 1 median IIC = .14; quartile 2 median IIC = .17; quartile 3 median IIC = .17; quartile 4 median IIC = .20). In general, there are significant differences between the length and complexity quartiles (Kruskal–Wallis test, p value $\leq .001$).

For the quartiles of sensitivity and social desirability, there are significant differences between the quartile IICs (Kruskal–Wallis test, p value $\leq .001$), but there were no apparent patterns for the characteristics of length and complexity/difficulty. For sensitivity, there is an increase in median IICs across the first three quartiles, followed by a decrease for the fourth quartile (quartile 1 median IIC = .18; quartile 2 median IIC = .18; quartile 3 median IIC = .22; quartile 4 median IIC = .13). For social desirability, there is a decrease in median IICs across the first three quartiles, followed by an increase for the fourth quartile (quartile 1 median IIC = .21; quartile 2 median IIC = .17; quartile 3 median IIC = .14; quartile 4 median IIC = .21).

Figure 7 IIC by quartiles of question characteristics



*** Significant at $p \leq .001$; ** significant at $.001 < p \leq .05$; * significant at $.05 < p \leq .1$; and not significant at $p > .1$ for the Mann–Whitney–Wilcoxon test

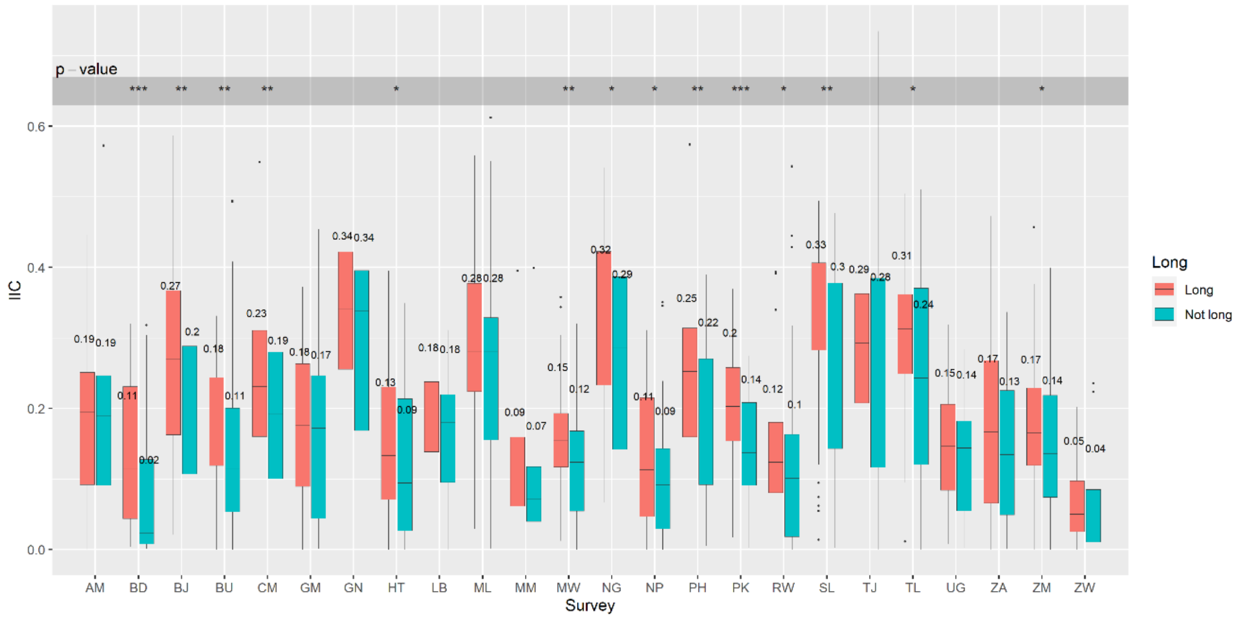
3.3 Intra-interviewer Correlations by Question Characteristics across Surveys

3.3.1 Uncontrolled comparisons

Figure 8 presents boxplots of IICs for question length across surveys. In general, across all surveys, the median IICs of long questions are higher than median IICs of shorter questions. These differences were found to be significant, in 14 of the 24 surveys at a significance level that is 10% or less (Mann–Whitney–

Wilcoxon test: p value ≤ .001 in 2 countries; .001 < p value ≤ .05 in 6 countries; .05 < p value ≤ .1 in 6 countries; p value > .1 in 10 countries).

Figure 8 IICs according to question length across surveys

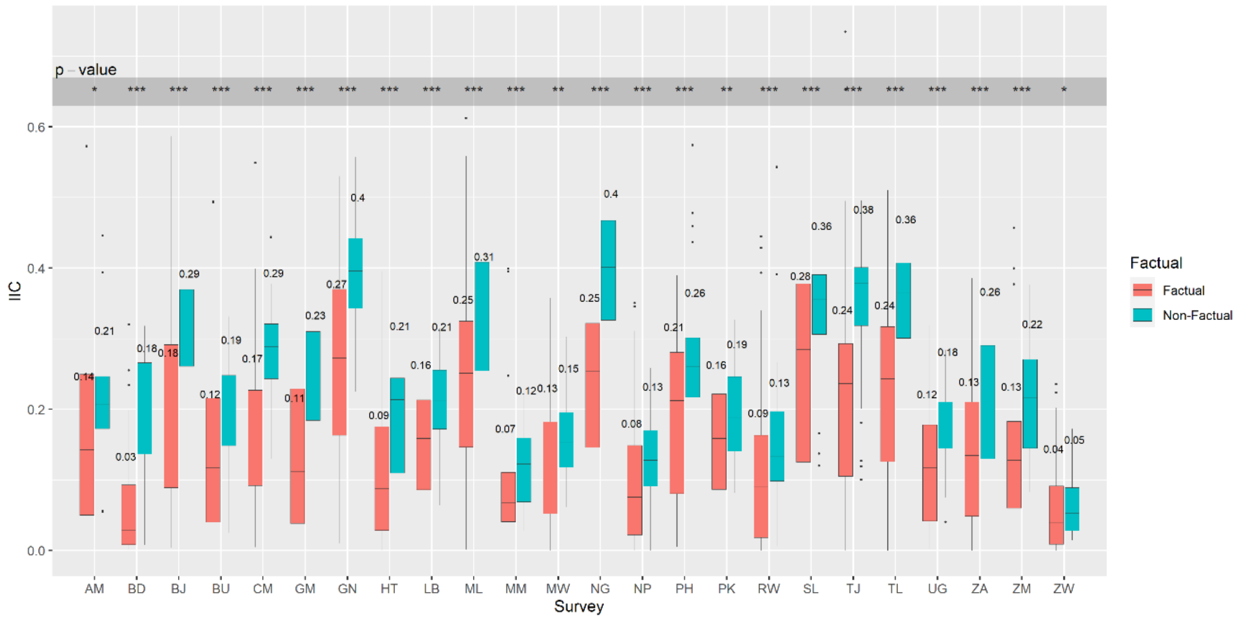


*** Significant at p ≤ .001; ** significant at .001 < p ≤ .05; * significant at .05 < p ≤ .1; and not significant at p > .1 for the Mann–Whitney–Wilcoxon test

Note: See page xv for a list of country codes used in this report.

Figure 9 presents boxplots of IICs according to question type (factual versus not factual) across surveys. Across all surveys, the median IICs of non-factual questions are significantly higher than the median IICs of factual questions at a significance level of 10% or less. For most surveys, the difference is statistically significant (Mann–Whitney–Wilcoxon test: p value ≤ .001 in 20 countries; .001 < p value ≤ .05 in 2 countries; .05 < p value ≤ .1 in 2 countries).

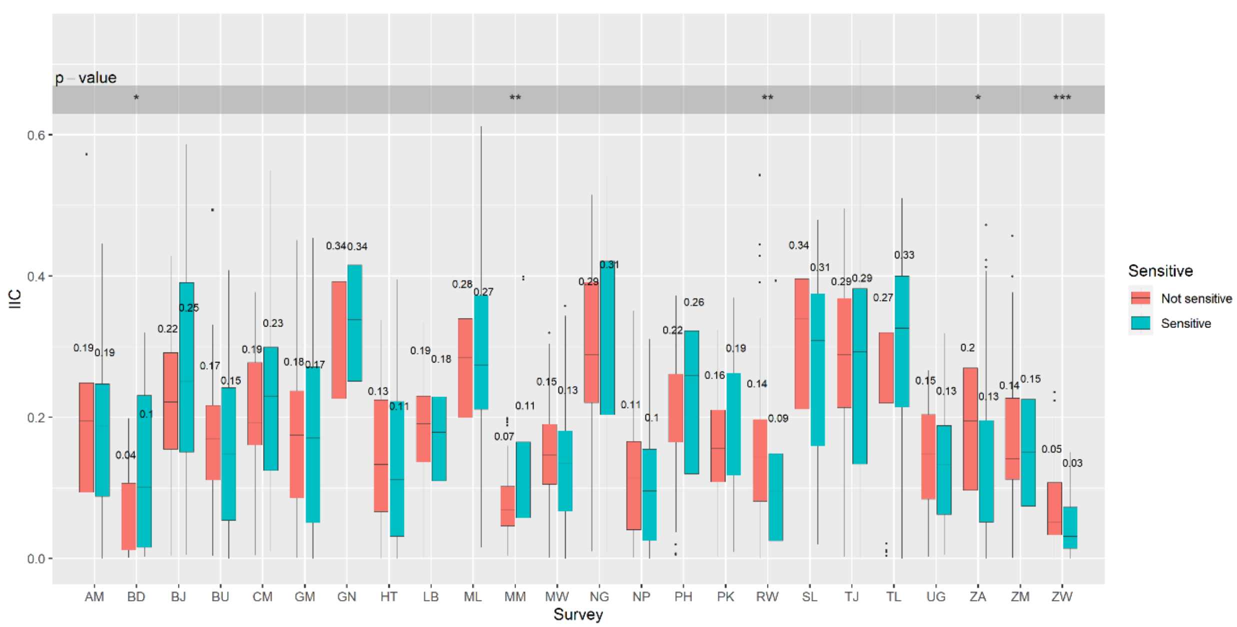
Figure 9 IICs according to question type across surveys



*** Significant at $p \leq .001$; ** significant at $.001 < p \leq .05$; * significant at $.05 < p \leq .1$; and not significant at $p > .1$ for the Mann–Whitney–Wilcoxon test
Note: See page xv for a list of country codes used in this report.

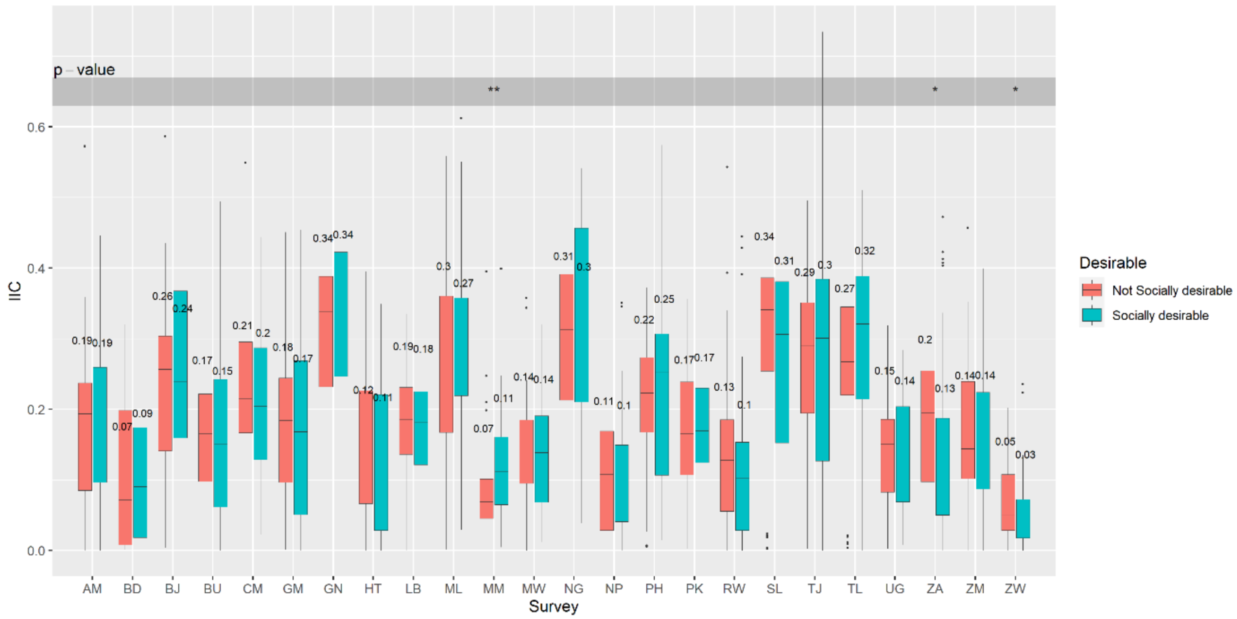
See page xv for a list of country codes used in this report.Figures 10 and 11 present boxplots of IICs according to question sensitivity and social desirability across surveys. We do not see a specific pattern for the median IICs by sensitivity or social desirability across surveys.

Figure 10 IICs according to question sensitivity across surveys



*** Significant at $p \leq .001$; ** significant at $.001 < p \leq .05$; * significant at $.05 < p \leq .1$; and not significant at $p > .1$ for the Mann–Whitney–Wilcoxon test
Note: See page xv for a list of country codes used in this report.

Figure 11 IICs according to social desirability across surveys

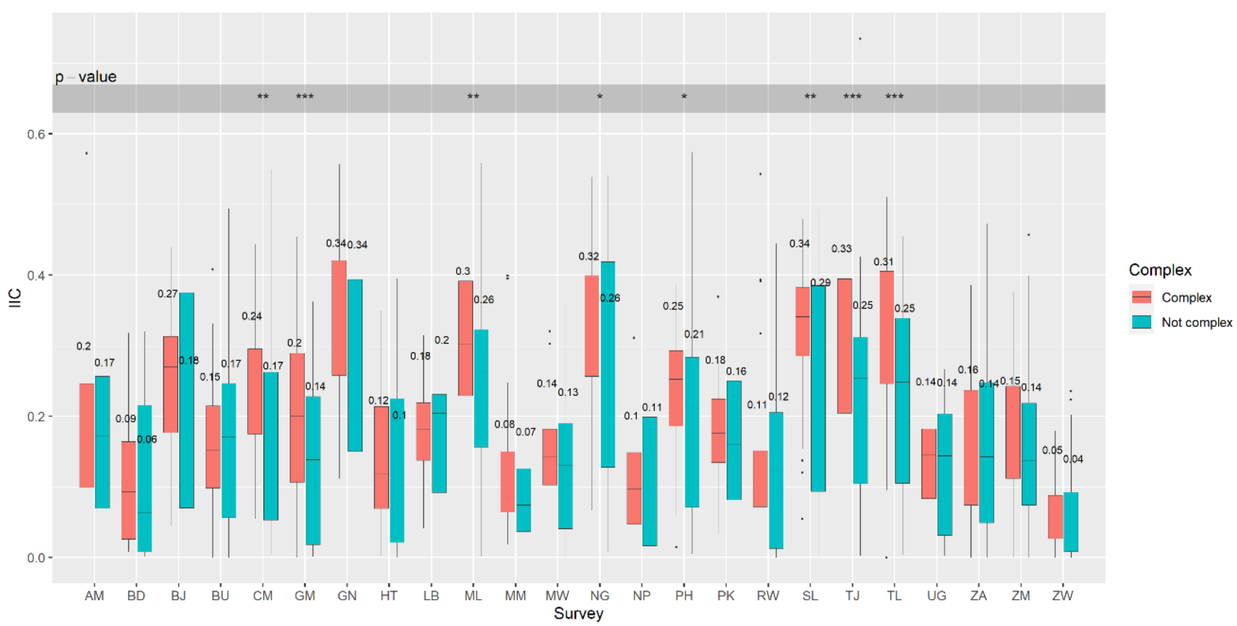


*** Significant at $p \leq .001$; ** significant at $.001 < p \leq .05$; * significant at $.05 < p \leq .1$; and not significant at $p > .1$ for the Mann–Whitney–Wilcoxon test

Note: See page xv for a list of country codes used in this report.

See page xv for a list of country codes used in this report. Figure 12 presents boxplots of IICs for question complexity/difficulty across surveys. In general, across most surveys, the median IICs of complex/difficult questions are higher than the median IICs of less complex/difficult questions. However, these differences were significant for only a few of the surveys (Mann–Whitney–Wilcoxon test: p value $\leq .001$ in 3 countries; $.001 < p$ value $\leq .05$ in 3 countries; $.05 < p$ value $\leq .1$ in 2 countries, p value $> .1$ in 16 countries).

Figure 12 IICs according to question type across surveys



*** Significant at $p \leq .001$; ** significant at $.001 < p \leq .05$; * significant at $.05 < p \leq .1$; and not significant at $p > .1$ for the Mann–Whitney–Wilcoxon test

Note: See page xv for a list of country codes used in this report.

3.3.2 Controlled comparisons

We used beta regression models to model IICs with question characteristics as model covariates. As indicated in Table 6, after controlling for other characteristics, in most countries the IICs of long questions and subjective (non-factual) questions remain significantly higher than IICs of shorter and factual questions, respectively. For length, in 17 countries the IICs of long questions were significantly higher than the IICs of shorter questions. For the question type, in 22 countries the IICs of non-factual questions were significantly higher than the IICs of factual questions. Similarly, after controlling for other characteristics, the IICs of complex or difficult questions were higher than the IICs of less complex questions in 20 countries, but only 6 were significant. Similar to the uncontrolled version, comparisons according to sensitivity or social desirability did not show any specific patterns or trends.

Table 6 Beta regression models of IICs: Estimated parameters and *p* values

Question characteristics	Long		Subjective		Sensitive		Socially desirable		Complex	
Survey	*β*	*p value*	*β*	*p value*	*β*	*p value*	*β*	*p value*	*β*	*p value*
Armenia 2015–16	0.504	0.006 *	0.468	0.026 *	0.100	0.668	0.182	0.432	0.215	0.304
Bangladesh 2017–18	0.316	0.189	1.225	0.000 *	0.125	0.654	-0.364	0.205	-0.167	0.471
Benin 2017–18	0.349	0.011 *	0.591	0.000 *	-0.027	0.878	0.014	0.939	0.223	0.139
Burundi 2016–17	0.177	0.275	1.095	0.000 *	-0.371	0.077	-0.150	0.470	-0.557	0.003 *
Cameroon 2018	0.348	0.004 *	0.696	0.000 *	0.136	0.393	-0.280	0.081	0.240	0.071
Gambia 2019–20	0.228	0.154	0.704	0.000 *	-0.222	0.295	-0.090	0.672	0.374	0.037 *
Guinea 2018	0.212	0.084	0.601	0.000 *	-0.031	0.845	-0.044	0.785	0.109	0.421
Haiti 2016–17	0.505	0.002 *	0.807	0.000 *	0.200	0.347	-0.322	0.132	0.102	0.586
Liberia 2019–20	0.260	0.042 *	0.488	0.001 *	-0.073	0.660	0.056	0.737	0.067	0.637
Malawi 2015–16	0.399	0.005 *	0.305	0.062	-0.614	0.001 *	0.538	0.004 *	0.398	0.011 *
Mali 2018	0.244	0.060	0.464	0.002 *	0.012	0.943	-0.035	0.838	0.246	0.087
Myanmar 2015–16	0.240	0.110	0.291	0.080	0.199	0.305	0.108	0.577	0.267	0.100
Nepal 2016	0.387	0.019 *	0.708	0.000 *	-0.604	0.007 *	0.315	0.149	0.090	0.624
Nigeria 2018	0.360	0.004 *	0.778	0.000 *	-0.147	0.376	0.032	0.845	0.118	0.396
Pakistan 2017–18	0.497	0.000 *	0.378	0.010 *	0.096	0.541	-0.120	0.451	0.030	0.833
Philippines 2017	0.300	0.025 *	0.508	0.002 *	0.066	0.702	-0.231	0.194	0.168	0.277
Rwanda 2019–20	0.583	0.001 *	1.011	0.000 *	-0.546	0.014 *	-0.209	0.345	-0.206	0.271
Sierra Leone 2019	0.389	0.004 *	0.444	0.004 *	-0.096	0.587	-0.145	0.416	0.481	0.001 *
South Africa 2016	0.162	0.385	1.027	0.000 *	-0.229	0.334	-0.405	0.084	-0.091	0.639
Tajikistan 2017	0.308	0.028 *	0.710	0.000 *	-0.263	0.149	-0.057	0.753	0.334	0.033 *
Timor-Leste 2016	0.290	0.029 *	0.477	0.001 *	0.066	0.692	-0.031	0.855	0.468	0.001 *
Uganda 2016	0.283	0.037 *	0.622	0.000 *	-0.179	0.321	-0.013	0.942	0.132	0.374
Zambia 2018	0.404	0.005 *	0.583	0.001 *	-0.123	0.529	0.060	0.758	0.165	0.311
Zimbabwe 2015	0.482	0.002 *	0.376	0.036 *	-0.446	0.031 *	-0.162	0.433	0.338	0.049 *
$\beta > 0/p \leq 0.05$	24	17	24	22	9	4	8	1	20	7

* Significant at $p \leq .05$

4 DISCUSSION

Although the DHS surveys have undergone many developments in data collection tools that have improved data quality, such as the growing use of computer-assisted personal interviewing (CAPI) systems, the interaction between interviewer and respondent remains crucial to understanding the quality of data collection. The interviewer's handling of the questions might vary according to question characteristics, and might lead to added variance in the data. In this report, we focused on the interviewer effect as defined by Kish (1962) as the increase in variance of sample statistics due to the interviewers. We used the Kish IIC, the interviewer intraclass correlation coefficient, as a measure of the interviewer effect.

We used multilevel models to estimate IICs on more than 100 DHS questions from 24 surveys. The questions covered a broad range of topics from the DHS Woman's Questionnaire. We modeled interviewer effects after controlling for respondent, interviewer, and sampling cluster characteristics. We examined the interviewer effects across countries and across different question characteristics, such as length, sensitivity, social desirability, complexity and/or difficulty, and question type (whether the information collected by the question was factual on non-factual). For defining some of the question characteristics such as sensitivity, social desirability, and complexity and/or difficulty, we used external raters.

Some question characteristics were shown to be associated with interviewer effects. Long questions, non-factual questions, and questions on complex or difficult topics were associated with larger interviewer effects compared to shorter questions, factual questions, and questions on less complex or difficult topics. These differences were consistent across most surveys, and with recent findings from the National Health Interview Survey (NHIS) (Dahlhamer et al. 2020). Moreover, the variable interviewer effects across countries are consistent with findings from the European Social Survey (ESS) (Beullens and Loosveldt 2016). In addition to the bivariate analysis, we examined the differences after controlling for other question characteristics with Beta regression to model IIC with question characteristics as covariates. The differences we noticed earlier remained after controlling for other question characteristics. Two question characteristics did not show any association with interviewer effects—those related to question sensitivity and social desirability.

There are some limitations of our study:

- In modeling the interviewer effects, we controlled for the available characteristics of interviewers across surveys, and some basic background characteristics of respondents, as well as geographic regions and residence type of sampling clusters (the only accessible data on the sampling cluster level). We realize that the models we used might not have adequately adjusted for other respondent and area effects, which might lead to over or under-estimation of interviewer effects.

- We did not include any questions from Sections 5 and 6, which measured child immunization, child health, and child nutrition. In these sections, detailed data about vaccinations and nutrition are collected about children under age 5. This involves collecting data recorded on vaccination cards or based on the mother's memory. We believe that these sections are worthy of a separate study to examine interviewer effect on their questions. We also considered the core Woman's Questionnaire only. Optional modules such as domestic violence were not considered in this study because we believe a separate study is needed to study the interviewer effects on the domestic violence questions.

- In defining the question characteristics, we used the English version of the DHS Woman's Questionnaire. We realize that characteristics such as length and complexity of questions might vary across countries, or even within the same country, according to the questionnaire language used for administration. In most countries, more than one version is used and administered according to the common language of the area. Therefore, the study did not account for the impact of questionnaire language on the question characteristics and on the interviewer effects determined in this analysis.

- In DHS surveys, interviewers are structured into teams that move together and are assigned to work in the same sampling clusters. We realize that this might contribute to the interviewer effects because team members might be quite similar in how they handle different questions. We believe a separate study is needed to study the structure of the interviewing teams and how that structure might contribute to interviewer effects.

In the future, we plan to conduct similar analyses on additional questions from different DHS questionnaires, and to further explore interviewer effects by respondent and interviewer characteristics. We are interested in knowing if:

- The results found here would be found for respondents from different backgrounds

- These results would be found for all interviewers from different backgrounds and with different experience levels

- These results hold within the same survey, especially when the interviewer was gaining more experience in asking the questions and building rapport with respondents

- The interviewer's workload plays any role in these results, especially with the recent findings from the European Social Survey (Wuyts and Loosveldt 2020)

Results from these studies can be used to identify question and interviewer characteristics with the largest IICs, which contribute significantly to variance. Moreover, results from the current study can be used to identify surveys with high and low interviewer effects. Future research can be done to investigate the differences between these countries in terms of the interviewer trainings conducted before the fieldwork, the monitoring of interviewers during fieldwork, and the interviewer workloads. This can help in developing the protocols used in these areas, so that the survey data are less affected by the interviewer effects.

REFERENCES

Amos, M. 2018. “Interviewer Effects on Patterns of Nonresponse: Evaluating the Impact on the Reasons for Contraceptive Nonuse in the Indonesia and the Philippines DHS.” *Demographic Research* 39 (14): 415–443. https://www.demographic-research.org/volumes/vol39/14/

Beullens, K., and G. Loosveldt. 2016. Interviewer Effects in the European Social Survey. *Survey Research Methods,* 10 (2): 103-118. https://doi.org/10.18148/srm/2016.v10i2.6261

Dahlhamer, J.M., B. Zablotsky, C. Zelaya, and A. Maitland. 2020. “Modeling Interviewer Effects in the National Health Interview Survey.” In *Interviewer Effects from a Total Survey Error Perspective*, edited by K. Olson, J.D. Smyth, J. Dykema, A.L. Holbrook, F. Kreuter, and B.T. West, 295-309. Boca Raton, FL, USA: CRC Press. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1055&context=sociw

Dahlhamer, J. M., A. Maitland, H. Ridolfo, A. Allen, and D. Brooks. 2019. “Exploring the Associations Between Question Characteristics, Respondent Characteristics, Interviewer Characteristics and Survey Data Quality.” In *Advances in Questionnaire Design, Development, Evaluation and Testing*, edited by P. Beatty, D. Collins, L. Kaye, J. L. Padilla, G. Willis, and A. Wilmot, 153-192. New York: Wiley.

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://doi.org/10.1093/her/cyp046

Footman, K. 2021. “Interviewer Effects on Abortion Reporting: A Multilevel Analysis of Household Survey Responses in Côte d’Ivoire, Nigeria and Rajasthan, India.” *BMJ Open* 11 (11): e047570. http://dx.doi.org/10.1136/bmjopen-2020-047570

Kish, L. 1962. “Studies of Interviewer Variance for Attitudinal Variables.” *Journal of the American Statistical Association* 57 (297): 92–115. https://doi.org/10.2307/2282442

Koo, T. K., and M.Y. Li. 2016. “Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research.” *Journal of Chiropractic Medicine* 15 (2): 155-63. https://doi.org/10.1016/j.jcm.2016.02.012

Leone, T., L. Sochas, and E. Coast. 2021. “Depends Who's Asking: Interviewer Effects in Demographic and Health Surveys Abortion Data.” *Demography* 58 (1): 31-50. https://doi.org/10.1215/00703370-8937468

Mangione, T. W., F. J. Fowler, Jr., and T. A. Louis. 1992. “Question Characteristics and Interviewer Effects.” *Journal of Official Statistics* 8 (3): 293–307. https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/question-characteristics-and-interviewer-effects.pdf

Pickery, J., and G. Loosveldt. 2001. “An Exploration of Question Characteristics that Mediate Interviewer Effects on Item Nonresponse.” *Journal of Official Statistics* 17 (3): 337–350. https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/an-exploration-of-question-characteristics-that-mediate-interviewer-effects-on-item-nonresponse.pdf

Olson, K., J.D. Smyth, J. Dykema, A. L.Holbrook, F. Kreuter, and B.T. 2020. *Interviewer Effects from a Total Survey Error Perspective* (1st ed.). New York: Chapman and Hall/CRC. https://doi.org/10.1201/9781003020219

Pullum, T. W., C. Juan, N. Khan, and S. Staveteig. 2018. *The Effect of Interviewer Characteristics on Data Quality in DHS Surveys.* DHS Methodological Reports No. 24. Rockville, Maryland, USA: ICF. https://dhsprogram.com/pubs/pdf/MR24/MR24.pdf

Schnell, R., and F. Kreuter. 2005. "Separating Interviewer and Sampling-Point Effects." *Journal of Official Statistics* 21 (3): 389–410.

Singh, A., K. Kumar, and F. Arnold. 2022. "How Interviewers Affect Responses to Sensitive Questions on the Justification for Wife Beating, the Refusal to have Conjugal Sex, and Domestic Violence in India." *Studies in Family Planning* 53 (2): 269-279. https://doi.org/10.1111/sifp.12193

Snijders, T., and R. Bosker. 1999. *Multilevel Analysis*. Thousand Oaks, CA, USA: London: Sage.

West, B. T., and A. Blom. 2017. "Explaining Interviewer Effects: A Research Synthesis." *Journal of Survey Statistics and Methodology* 5 (2): 175–211. https://doi.org/10.1093/jssam/smw024

Wuyts, C. and G. Loosveldt. 2020. "Measurement of Interviewer Workload within the Survey and an Exploration of Workload Effects on Interviewers' Field Efforts and Performance." *Journal of Official Statistics* 36 (3): 561-588. https://doi.org/10.2478/jos-2020-0029

APPENDIX: QUESTIONS, VARIABLES, AND OUTCOMES BY CHARACTERISTICS

Appendix Table A.1 Questions, variables, and outcomes with classification by question characteristics

Question no.	Question	Variables	Category	Outcome	Factual	Quartiles: Length	Sensitivity	Social desirability	Difficulty/ complexity
113	Do you read a newspaper or magazine at least once a week, less than once a week or not at all?	v157	0 1 2	read_no read_less read_once	Yes	2	2	3	1
114	Do you listen to the radio at least once a week, less than once a week or not at all?	v158	0 1 2	listen_no listen_less listen_once	Yes	2	1	2	1
115	Do you watch television at least once a week, less than once a week or not at all?	v159	0 1 2	watch_no watch_less watch_once	Yes	2	1	2	1
116	Do you own a mobile phone?	v169a	1	own_phone	Yes	1	1	3	1
117	Do you use your mobile phone for any financial transactions?	v169b	1	phone_transact	Yes	1	3	3	2
118	Do you have an account in a bank or other financial institution that you yourself use?	v170	1	own_account	Yes	2	3	2	3
119 120	Have you ever used the Internet? In the last 12 months, have you used the Internet? IF NECESSARY, PROBE FOR USE FROM ANY LOCATION, WITH ANY DEVICE.	in v171a	1 2	internet_inyear internet_beforeyear	Yes	3	1	2	3
121	During the last one month, how often did you use the internet: almost every day, at least once a week, less than once a week, or not at all?	v171b	0 1 2 3	internet_notatall internet_lessweek internet_leastweek internet_daily	Yes	3	1	2	3
203a	How many sons live with you?	v202	0 1 2 3	sonshome_0 sonshome_1 sonshome_2 sonshome_3plus	Yes	1	1	1	1
203b	And how many daughters live with you?	v203	0 1 2 3	daughtershome_0 daughtershome_1 daughtershome_2 daughtershome_3plus	Yes	1	1	1	1
205a	How many sons are alive but do not live with you?	v204	0 1 2	sonselse_0 sonselse_1 sonselse_2plus	Yes	1	2	2	1
205b	And how many daughters are alive but do not live with you?	v205	0 1 2	daughterselse_0 daughterselse_1 daughterselse_2plus	Yes	1	3	2	1

(continued…)

Appendix Table A.1—*Continued*

Question no.	Question	Variables	Category	Outcome	Factual	Quartiles: Length	Quartiles: Sensitivity	Quartiles: Social desirability	Quartiles: Difficulty/ complexity
207a	How many boys have died?	v206	0 1 2	sonsdied_0 sonsdied_1 sonsdied_2plus	Yes	1	4	3	1
207b	And how many girls have died?	v207	0 1 2	daughtersdied_0 daughtersdied_1 daughtersdied_2plus	Yes	1	4	3	1
226	Are you pregnant now?	v213	1	current_pregnant	Yes	1	4	3	1
228	When you got pregnant, did you want to get pregnant at that time?	in v225	1 2 3	currpregnancy_wanted currpregnancy_later currpregnancy_notatall	Yes	1	4	4	3
230	Have you ever had a pregnancy that miscarried, was aborted, or ended in a stillbirth?	v228	1	terminate_pregnancy	Yes	2	4	4	4
240 241	From one menstrual period to the next, are there certain days when a woman is more likely to become pregnant? Is this time just before her period begins, during her period, right after her period has ended, or halfway between two periods?	in v217	0 1 2 3 4 5	period_during period_after period_middle period_before period_anytime period_otherdk	No	4	1	1	4
242	After the birth of a child, can a woman become pregnant before her menstrual period has returned?	v244	0 1 2	pregnantreturn_no pregnantreturn_yes pregnantreturn_dk	No	2	1	1	3
301-01	Now I would like to talk about family planning - the various ways or methods that a couple can use to delay or avoid a pregnancy. Have you ever heard of Female Sterilization? PROBE: Women can have an operation to avoid having any more children.	v304_01	1	knowmethod_fsterilize	Yes	4	1	3	1
301-02	Now I would like to talk about family planning - the various ways or methods that a couple can use to delay or avoid a pregnancy. Have you ever heard of Male Sterilization? PROBE: Men can have an operation to avoid having any more children.	v304_02	1	knowmethod_msterilize	Yes	4	1	3	2
301-03	Now I would like to talk about family planning - the various ways or methods that a couple can use to delay or avoid a pregnancy. Have you ever heard of the use of an IUD? PROBE: Women can have a loop or coil placed inside them by a doctor or a nurse which can prevent pregnancy for one or more years.	v304_03	1	knowmethod_iud	Yes	4	1	3	1

(continued…)

Appendix Table A.1—*Continued*

Question no.	Question	Variables	Category	Outcome	Factual	Quartiles: Length	Quartiles: Sensitivity	Quartiles: Social desirability	Quartiles: Difficulty/ complexity
301-04	Now I would like to talk about family planning - the various ways or methods that a couple can use to delay or avoid a pregnancy. Have you ever heard of the use of Injectables? PROBE: Women can have an injection by a health provider that stops them from becoming pregnant for one or more months.	v304_04	1	knowmethod_injectable	Yes	4	1	3	1
301-05	Now I would like to talk about family planning - the various ways or methods that a couple can use to delay or avoid a pregnancy. Have you ever heard of the use of Implants? PROBE: Women can have one or more small rods placed in their upper arm by a doctor or nurse which can prevent pregnancy for one or more years.	v304_05	1	knowmethod_implant	Yes	4	1	3	1
301-06	Now I would like to talk about family planning - the various ways or methods that a couple can use to delay or avoid a pregnancy. Have you ever heard of the use of a Pill? PROBE: Women can take a pill every day to avoid becoming pregnant.	v304_06	1	knowmethod_pill	Yes	4	1	3	1
301-07	Now I would like to talk about family planning - the various ways or methods that a couple can use to delay or avoid a pregnancy. Have you ever heard of the use of a Condom? PROBE: Men can put a rubber sheath on their penis before sexual intercourse.	v304_07	1	knowmethod_mcondom	Yes	4	2	3	1
301-08	Now I would like to talk about family planning - the various ways or methods that a couple can use to delay or avoid a pregnancy. Have you ever heard of the use of a Female Condom? PROBE: Women can place a sheath in their vagina before sexual intercourse.	v304_08	1	knowmethod_fcondom	Yes	4	2	3	2
301-09	Now I would like to talk about family planning - the various ways or methods that a couple can use to delay or avoid a pregnancy. Have you ever heard of the use of a Emergency Contraception? PROBE: As an emergency measure, within 3 days after they have unprotected sexual intercourse, women can take special pills to prevent pregnancy.	v304_09	1	knowmethod_specialpill	Yes	4	1	3	2

(continued…)

Appendix Table A.1—*Continued*

Question no.	Question	Variables	Category	Outcome	Factual	Quartiles Length	Sensitivity	Social desirability	Difficulty/ complexity
301-10	Now I would like to talk about family planning - the various ways or methods that a couple can use to delay or avoid a pregnancy. Have you ever heard of the use of the Standard Days Method? PROBE: A woman uses a string of colored beads to know the days she can get pregnant. On the days she can get pregnant, she uses a condom or does not have sexual intercourse.	v304_10	1	knowmethod_stddays	Yes	4	1	3	2
301-11	Now I would like to talk about family planning - the various ways or methods that a couple can use to delay or avoid a pregnancy. Have you ever heard of the use of the Lactational Amenorrhea Method (LAM)? PROBE: Up to 6 months after childbirth, before the menstrual period has returned, women use a method requiring frequent breastfeeding day and night.	v304_11	1	knowmethod_lam	Yes	4	1	3	3
301-12	Now I would like to talk about family planning - the various ways or methods that a couple can use to delay or avoid a pregnancy. Have you ever heard of the use of the Rhythm Method? PROBE: To avoid pregnancy, women do not have sexual intercourse on the days of the month they think they can get pregnant.	v304_12	1	knowmethod_rhythm	Yes	4	1	3	2
301-13	Now I would like to talk about family planning - the various ways or methods that a couple can use to delay or avoid a pregnancy. Have you ever heard of the use of the Withdrawal Method? PROBE: Men can be careful and pull out before climax.	v304_13	1	knowmethod_withdraw	Yes	4	3	3	2
301-14	Have you heard of any other ways or methods that women or men can use to avoid pregnancy?	v304_14	1	knowmethod_other	Yes	2	1	3	3
303	Are you or your partner currently doing something or using any method to delay or avoid getting pregnant?	in v312			Yes	3	4	3	2
314	Have you ever used anything or tried in any way to delay or avoid getting pregnant?	v302a	0 1 2	usedcontraceptive_no usedcontraceptive_ yesnocalendar usedcontraceptive_ yescalendar	Yes	2	4	3	2
327	In the last 12 months, were you visited by a fieldworker?	v393	1	visitfieldworker12m	Yes	1	1	1	2
329	In the last 12 months, have you visited a health facility for care for yourself or your children?	v394	1	visitfacility12m	Yes	2	2	1	2

(continued…)

Appendix Table A.1—*Continued*

Question no.	Question	Variables	Category	Outcome	Factual	Quartiles: Length	Sensitivity	Social desirability	Difficulty/ complexity
412	How many times did you receive antenatal care during this pregnancy?	m14	0 1 2 3	ancare_no_dk ancare_1_4 ancare_5_9 ancare_10_more	Yes	1	3	3	4
413a	As part of your antenatal care during this pregnancy, were any of the following done at least once: Was your blood pressure measured?	m42c	1	ancare_bp	Yes	3	2	2	4
413b	As part of your antenatal care during this pregnancy, were any of the following done at least once: Did you give a urine sample?	m42d	1	ancare_urine	Yes	3	2	2	4
413c	As part of your antenatal care during this pregnancy, were any of the following done at least once: Did you give a blood sample?	m42e	1	ancare_blood	Yes	3	2	2	4
414 415	During this pregnancy, were you given an injection in the arm to prevent the baby from getting tetanus, that is, convulsions after birth? During this pregnancy, how many times did you get a tetanus injection?	in m1	0 1 2	tetanus_no_dk tetanus_1 tetanus_2plus	Yes	4	2	2	4
417 418	At any time before this pregnancy, did you receive any tetanus injections? Before this pregnancy, how many times did you receive a tetanus injection?	in m1a	0 1 2	tetanusbefore_no_dk tetanusbefore_1 tetanusbefore_2plus	Yes	4	2	2	4
420	During this pregnancy, were you given or did you buy any iron tablets or iron syrup? SHOW TABLETS/SYRUP.	m45	1	given_iron	Yes	3	2	3	4
422	During this pregnancy, did you take any drug for intestinal worms?	m60	1	drug_worm	Yes	1	3	2	4
426	When (NAME) was born, was (NAME) very large, larger than average, average, smaller than average, or very small?	m18	1 2 3 4 5 6	childsize_verylarge childsize_large childsize_average childsize_small childsize_verysmall childsize_dk	No	2	3	3	2
427	Was (NAME) weighed at birth?	m19a	0 1 2	weighedatbirth_no_dk_sp weighedatbirth_card weighedatbirth_recall	Yes	1	2	1	2
432	Was (NAME) delivered by caesarean, that is, did they cut your belly open to take the baby out?	m17	1	caesarean_yes	Yes	2	3	1	1

(continued…)

Appendix Table A.1—*Continued*

Question no.	Question	Variables	Category	Outcome	Factual	Quartiles: Length	Sensitivity	Social desirability	Difficulty/ complexity
433	When was the decision made to have the caesarean section? Was it before or after your labor pains started?	m17a	1	caesarean_decision	Yes	3	3	2	4
434 434A	Immediately after the birth, was (NAME) put on your chest? Was (NAME)'s bare skin touching your bare skin?	in m77	0 1	child_motherstouch_no_dk child_motherstouch_yes	Yes	2	2	2	4
435	I would like to talk to you about checks on your health after delivery, for example, someone asking you questions about your health or examining you. Did anyone check on your health while you were still in the facility?	m62	1	healthcheck_yes	Yes	4	3	2	4
709	Have you been married or lived with a man only once or more than once?	v503	1	union_2plus	Yes	1	4	3	1
713	Now I would like to ask some questions about sexual activity in order to gain a better understanding of some important life issues. Let me assure you again that your answers are completely confidential and will not be told to anyone. If we should come to any question that you don't want to answer, just let me know and we will go to the next question. How old were you when you had sexual intercourse for the very first time?	v525	1 2 3 4	agefirstsex_no_firstunion_dk agefirstsex_15_19 agefirstsex_20_24 agefirstsex_25plus	Yes	4	4	4	3
723	In total, with how many different people have you had sexual intercourse in the last 12 months?	in v766b	0 1 2	partner_0 partner_1_dk partner_2plus	Yes	2	4	4	2
727	In total, with how many different people have you had sexual intercourse in your lifetime?	v836	1 2 3 4	lifetimepartner_1_dk lifetimepartner_2 lifetimepartner_3 lifetimepartner_4plus	Yes	2	4	4	3
804	Now I have some questions about the future. Would you like to have (a/another) child, or would you prefer not to have any (more) children?	in v602	1 2 3 4	havechild_another havechild_undecided havechild_nomore havechild_sterilized_infecunt	No	3	3	3	3
812	Do you think you will use a contraceptive method to delay or avoid pregnancy at any time in the future?	v364	1 2 3 4	contraceptiveuse_modern contraceptiveuse_traditional contraceptiveuse_intent contraceptiveuse_nonuser	No	3	4	3	4

(continued…)

Appendix Table A.1—*Continued*

Question no.	Question	Variables	Category	Outcome	Factual	Quartiles: Length	Sensitivity	Social desirability	Difficulty/ complexity
814	How many of these children would you like to be boys, how many would you like to be girls and for how many would it not matter if it's a boy or a girl?	v627: v629	0 1 2 3 0 1 2 3	preferboys_0 preferboys_1 preferboys_2 preferboys_3plus prefergirls_0 prefergirls_1 prefergirls_2 prefergirls_3plus	No	3	3	3	3
815a	In the last few months have you heard about family planning on the radio?	v384a	1	fplan_radio	Yes	2	1	1	3
815b	In the last few months have you seen anything about family planning on the television?	v384b	1	fplan_tv	Yes	2	1	1	3
815c	In the last few months have you read about family planning in a newspaper or magazine?	v384c	1	fplan_paper	Yes	2	1	1	3
815d	In the last few months have you received a voice or text message about family planning on a mobile phone?	v384d	1	fplan_phone	Yes	3	1	1	2
			1 2 3	usedecision_partner usedecision_joint usedecision_other					
822	Does your (husband/partner) want the same number of children that you want, or does he want more or fewer than you want?	v621	0 1 2 3	childwant_same childwant_more childwant_less childwant_dk	Yes	3	4	4	4
903 904	Did your (husband/partner) ever attend school? What was the highest level of school he attended: primary, secondary, or higher?	in v701	1 2 3 4	partnerschool_no_dk partnerschool_primary partnerschool_secondary partnerschool_higher	Yes	3	2	3	1
906 907	Has your (husband/partner) done any work in the last 7 days? Has your (husband/partner) done any work in the last 12 months?	in v704a	1 2 3	partnerwork_no_dk partnerwork_7days partnerwork_12months	Yes	3	3	3	1
909	Aside from your own housework, have you done any work in the last seven days?	in v714-v731	1	work_yes	Yes	2	2	3	2

(continued…)

Appendix Table A.1—*Continued*

						Quartiles			
Question no.	Question	Variables	Category	Outcome	Factual	Length	Sensitivity	Social desirability	Difficulty/ complexity
910	As you know, some women take up jobs for which they are paid in cash or kind. Others sell things, have a small business or work on the family farm or in the family business. In the last seven days, have you done any of these things or any other work?	v714	1	work_yes	Yes	4	2	3	2
912	Have you done any work in the last 12 months?	in v731	0 1 2	work12m_no work12m_pastyear work12m_current	Yes	1	2	3	2
919	Who usually decides how the money you earn will be used: you, your (husband/partner), or you and your (husband/partner) jointly?	v739	1 2 3	spenddecision_resp spenddecision_joint spenddecision_partner_ other	No	3	3	4	4
920	Would you say that the money that you earn is more than what your (husband/partner) earns, less than what he earns, or about the same?	v746	1 2 3 4	respearn_more respearn_less respearn_same respearn_no_dk	Yes	3	4	3	4
921	Who usually decides how your (husband's/partner's) earnings will be used: you, your (husband/partner), or you and your (husband/partner) jointly?	v743f	1 2 3 4	spenddecisionpart_resp spenddecisionpart_joint spenddecisionpart_ partner spenddecisionpart_noearn _other	No	4	4	4	4
922	Who usually makes decisions about health care for yourself: you, your (husband/partner), you and your (husband/partner) jointly, or someone else?	v743a	1 2 3 4	resphealth_alone resphealth_joint resphealth_partner resphealth_other	No	4	4	3	4
923	Who usually makes decisions about making major household purchases?	v743b	1 2 3 4	hhpurchases_alone hhpurchases_joint hhpurchases_partner hhpurchases_other	No	1	4	3	4
924	Who usually makes decisions about visits to your family or relatives?	v743d	1 2 3 4	familyvisits_alone familyvisits_joint familyvisits_partner familyvisits_other	No	1	3	3	3
925	Do you own this or any other house either alone or jointly with someone else?	v745a	1 2 3 4	ownhouse_no ownhouse_alone ownhouse_joint ownhouse_both	Yes	2	3	3	3

(continued…)

Appendix Table A.1—*Continued*

Question no.	Question	Variables	Category	Outcome	Factual	Quartiles: Length	Sensitivity	Social desirability	Difficulty/ complexity
932a	In your opinion, is a husband justified in hitting or beating his wife in the following situation: If she goes out without telling him?	v744a	1 2 3	beat_tell_yes beat_tell_no beat_tell_dk	No	3	3	4	2
932b	In your opinion, is a husband justified in hitting or beating his wife in the following situation: If she neglects the children?	v744b	1 2 3	beat_neglect_yes beat_neglect_no beat_neglect_dk	No	3	3	4	2
932c	In your opinion, is a husband justified in hitting or beating his wife in the following situations: If she argues with him?	v744c	1 2 3	beat_argue_yes beat_argue_no beat_argue_dk	No	3	3	4	2
932d	In your opinion, is a husband justified in hitting or beating his wife in the following situation: If she refuses to have sex with him?	v744d	1 2 3	beat_sex_yes beat_sex_no beat_sex_dk	No	3	3	4	2
932e	In your opinion, is a husband justified in hitting or beating his wife in the following situation: If she burns the food?	v744e	1 2 3	beat_food_yes beat_food_no beat_food_dk	No	3	3	4	2
1001	Now I would like to talk about something else. Have you ever heard of HIV or AIDS?	in v750-v751	1	heard_AIDS_no	Yes	2	2	3	1
1002	HIV is the virus that can lead to AIDS. Can people reduce their chance of getting HIV by having just one uninfected sex partner who has no other sex partners?	v754dp	1 2 3	hivriskreduce_no hivriskreduce_yes hivriskreduce_dk	No	4	2	3	4
1003	Can people get HIV from mosquito bites?	v754jp	1 2 3	mosquito_hiv_no mosquito_hiv_yes mosquito_hiv_dk	No	1	2	3	2
1004	Can people reduce their chance of getting HIV by using a condom every time they have sex?	v754cp	1 2 3	condom_hiv_no condom_hiv_yes condom_hiv_dk	No	2	2	3	3
1005	Can people get HIV by sharing food with a person who has HIV?	v754wp	1 2 3	sharingfood_hiv_no sharingfood_hiv_yes sharingfood_hiv_dk	No	1	2	3	2
1006	Can people get HIV because of witchcraft or other supernatural means?	v823	1 2 3	witchcraft_hiv_no witchcraft_hiv_yes witchcraft_hiv_dk	No	1	2	3	2
1007	Is it possible for a healthy-looking person to have HIV?	v756	1 2 3	healthyperson_hiv_no healthyperson_hiv_yes healthyperson_hiv_dk	No	1	1	3	2

(continued…)

Appendix Table A.1—*Continued*

Question no.	Question	Variables	Category	Outcome	Factual	Quartiles: Length	Sensitivity	Social desirability	Difficulty/ complexity
1008a	Can HIV be transmitted from a mother to her baby during pregnancy?	v774a			No	1	2	2	3
1008b	Can HIV be transmitted from a mother to her baby during delivery?	v774b			No	1	2	2	3
1008c	Can HIV be transmitted from a mother to her baby by breastfeeding?	v774c			No	1	1	2	3
1010	Are there any special drugs that a doctor or a nurse can give to a woman infected with HIV to reduce the risk of transmission to the baby?	v824	1 2 3	drugs_hivtransmission_no drugs_hivtransmission_yes drugs_hivtransmission_dk	No	3	2	2	3
1027	I don't want to know the results, but have you ever been tested for HIV?	v781	1	tested_hiv	Yes	1	4	4	2
1031	Do you know of a place where people can go to get an HIV test?	v783	1	hivtest_place	Yes	1	3	3	1
1033 1034	Have you heard of test kits people can use to test themselves for HIV? Have you ever tested yourself for HIV using a self-test kit?	in v856	1 2 3	hivkit_neverheard hivkit_heardandtested hivkit_heardbutnottested	Yes	3	3	3	2
1035	Would you buy fresh vegetables from a shopkeeper or vendor if you knew that this person had HIV?	v825	1 2 3	buyveg_hiv_no buyveg_hiv_yes buyveg_hiv_dk	No	2	4	4	3
1036	Do you think children living with HIV should be allowed to attend school with children who do not have HIV?	v857a	1 2 3	school_hiv_no school_hiv_yes school_hiv_dk	No	3	4	4	3
1037	Do you think people hesitate to take an HIV test because they are afraid of how other people will react if the test result is positive for HIV?	v857b	1 2 3	test_hiv_no test_hiv_yes test_hiv_dk	No	3	3	4	4
1038	Do people talk badly about people living with HIV, or who are thought to be living with HIV?	v857c	1 2 3	talkbad_hiv_no talkbad_hiv_yes talkbad_hiv_dk	No	2	3	4	3
1039	Do people living with HIV, or thought to be living with HIV, lose the respect of other people?	v857d	1 2 3	respect_hiv_no respect_hiv_yes respect_hiv_dk	No	2	3	4	4
1040	Do you agree or disagree with the following statement? I would be ashamed if someone in my family had HIV.	v777a	1 2 3	familyhiv_ashamed_disagree familyhiv_ashamed_agree familyhiv_ashamed_dk	No	3	4	4	3

(continued…)

Appendix Table A.1—*Continued*

Question no.	Question	Variables	Category	Outcome	Factual	Quartiles: Length	Quartiles: Sensitivity	Quartiles: Social desirability	Quartiles: Difficulty/ complexity
1041	Do you fear that you could get HIV if you come into contact with the saliva of a person living with HIV?	v858	1	hiv_saliva_no	No	2	4	4	4
			2	hiv_saliva_yes					
			3	hiv_saliva_dk					
1045	Now I would like to ask you some questions about your health in the last 12 months. During the last 12 months, have you had a disease which you got through sexual contact?	v763a	1	sti_12months_no	Yes	4	4	4	4
			2	sti_12months_yes					
			3	sti_12months_dk					
1046	Sometimes women experience a bad-smelling abnormal genital discharge. During the last 12 months, have you had a bad-smelling abnormal genital discharge?	v763c	1	discharge_12months_no	Yes	4	4	4	4
			2	discharge_12months_yes					
			3	discharge_12months_dk					
1047	Sometimes women have a genital sore or ulcer. During the last 12 months, have you had a genital sore or ulcer?	v763b	1	ulcer_12months_no	Yes	3	4	4	4
			2	ulcer_12months_yes					
			3	ulcer_12months_dk					
1051	If a wife knows her husband has a disease that she can get during sexual intercourse, is she justified in asking that they use a condom when they have sex?	v822	1	husband_usecondom_no	No	4	3	4	4
			2	husband_usecondom_yes					
			3	husband_usecondom_dk					
1052	Is a wife justified in refusing to have sex with her husband when she knows he has sex with other women?	v633b	1	husband_sexotherwomen_no	No	3	3	4	4
			2	husband_sexotherwomen_yes					
			3	husband_sexotherwomen_dk					
1054	Can you say no to your (husband/partner) if you do not want to have sexual intercourse?	v850a	1	husband_refusesex_no	No	2	4	4	4
			2	husband_refusesex_yes					
			3	husband_refusesex_dk					
1055	Could you ask your (husband/partner) to use a condom if you wanted him to?	v850b	1	husband_askusecondom_no	No	2	4	4	4
			2	husband_askusecondom_yes					
			3	husband_askusecondom_dk					
1104	Do you currently smoke cigarettes every day, some days, or not at all?	v463a	1	currentcigarettesmoker_yes	Yes	1	4	4	1
1106	Do you currently smoke or use any other type of tobacco every day, some days, or not at all?	in v463a:v463x	1	currentothersmoker_yes	Yes	2	4	4	1

(continued…)

Appendix Table A.1—*Continued*

Question no.	Question	Variables	Category	Outcome	Factual	Quartiles: Length	Sensitivity	Social desirability	Difficulty/ complexity
1108a	Many different factors can prevent women from getting medical advice or treatment for themselves. When you are sick and want to get medical advice or treatment, is each of the following a big problem or not a big problem: Getting permission to go to the doctor?	v467b	1	permission_doctor	Yes	4	3	1	3
1108b	Many different factors can prevent women from getting medical advice or treatment for themselves. When you are sick and want to get medical advice or treatment, is each of the following a big problem or not a big problem: Getting money needed for advice or treatment?	v467c	1	money_treatment	Yes	4	4	1	2
1108c	Many different factors can prevent women from getting medical advice or treatment for themselves. When you are sick and want to get medical advice or treatment, is each of the following a big problem or not a big problem: The distance to the health facility?	v467d	1	distance_healthfacility	Yes	4	3	1	2
1108d	Many different factors can prevent women from getting medical advice or treatment for themselves. When you are sick and want to get medical advice or treatment, is each of the following a big problem or not a big problem: Not wanting to go alone?	v467f	1	goalone_treatment	Yes	4	3	1	2
1109	Are you covered by any health insurance?	in v481a-v481x	1	healthinsurance_yes	Yes	1	3	1	2