USAID
FROM THE AMERICAN PEOPLE

DATA QUALITY IN DEMOGRAPHIC AND HEALTH SURVEYS THAT USED LONG AND SHORT QUESTIONNAIRES

DHS METHODOLOGICAL REPORTS 30

SEPTEMBER 2020

This publication was produced for review by the United States Agency for International Development. It was prepared by Courtney K. Allen, Julia Fleuret, and Jehan Ahmed.

DHS Methodological Reports No. 30

Data Quality in Demographic and Health Surveys That Used Long and Short Questionnaires

Courtney K. Allen[1]
Julia Fleuret[1]
Jehan Ahmed[1]

ICF
Rockville, Maryland, USA

September 2020

[1] The DHS Program, ICF

Corresponding author: Courtney K. Allen, International Health and Development, ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA; phone: +1 301-407-6500; fax: +1 301-407-6501; email: courtney.allen@icf.com

Acknowledgments: The authors extend their appreciation to many individuals who provided valuable input at different points in this study's development. Dr. Rukundo K. Benedict (The DHS Program) and Joy Fishel (The DHS Program) provided useful feedback on the key informant interview guide. Dr. Shireen Assaf (The DHS Program), Dr. Nicole Bellows (Avenir Health), and Trevor Croft (The DHS Program) provided thoughtful comments on an earlier draft of this study, and Dr. Shireen Assaf also provided thoughts on the quantitative analysis approach. Anjushree Pradhan (The DHS Program), Trevor Croft (The DHS Program), Dr. Fred Arnold (The DHS Program), and Dr. Gulnara Semenov (The DHS Program) gave input into the selection of surveys. Dr. Ruilin Ren (The DHS Program) provided clarifications and expertise on how survey samples are designed for our background section. We are also grateful to Dr. Tom Pullum (The DHS Program) for his input on and helpful discussion regarding the analysis and indicators included in this study. The key informants generously shared their time, expertise, and opinions; talking with them was not only informative but delightful. Finally, there would not be surveys to analyze without the dedication and skill of fieldworkers; our sincere thanks to the approximately 915 interviewers who worked on the 2016 South Africa Demographic and Health Survey, the 2014 Kenya Demographic and Health Survey, and the 2015-16 India National Family Health Survey.

Editor: Cathy Johnson
Document Production: Natalie Shattuck

This study was implemented with support provided by 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.

Recommended citation:

Allen, Courtney K., Julia Fleuret, Jehan Ahmed. 2020. *Data Quality in Demographic and Health Surveys That Used Long and Short Questionnaires*. DHS Methodological Reports No. 30. 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 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

Since the first phase of The DHS Program in 1984, DHS surveys have increased in scope and complexity as questionnaires are lengthened and survey modules are added. Increase in questionnaire length can increase interview time, and ultimately may present more burden for both the interviewer and the respondent. It seems intuitive that longer questionnaires would have different effects than shorter questionnaires on fieldwork, interviewer fatigue and performance, and survey implementation. Surveys that have two different lengths of questionnaires offer an opportunity to explore the extent to which questionnaires of different lengths may have these effects. This report aims to understand the effect, if any, of questionnaire length on data quality using the 2016 South Africa DHS, 2014 Kenya DHS, and 2015-16 India National Family Health Survey.

We described the differences in fieldwork and interview length between the long and short questionnaires and examine data quality indicators to see whether the different length of questionnaires experience differing data quality. We used two types of data quality indicators: indicators that may reflect efforts on the part of fieldworkers to reduce survey burden (i.e., their workload), and those concerning age and date of birth that are notoriously difficult to collect accurately in household surveys. Additionally, we explored themes resulting from qualitative interviews with survey experts who worked on surveys with long and short questionnaires to understand how these surveys are implemented, and data quality considerations, if any, along the survey process.

The long questionnaires in each country had large differences in the average number of variables per woman and interview length compared to short questionnaires. In Kenya, the long questionnaire lasted twice as long as the short questionnaire on average. Despite these differences, there is no evidence that interviewers may have intentionally reduced their workload. We found little evidence that having differing lengths of questionnaires resulted in data quality differences.

Key informants agree that deploying long and short questionnaires to obtain estimates of a subset of survey indicators for nonstandard populations or at lower administrative levels solves a significant problem in survey design and implementation. While challenges remain, it is widely embraced as a useful approach to survey design. Key informants were unanimously supportive of using long and short questionnaires in the future. We recommend that the use of long and short questionnaires be included in the survey design options for future DHS surveys so that surveys can meet in-country data demands while maintaining feasibility.

KEY WORDS: household survey, data quality, survey design

ACRONYMS AND ABBREVIATIONS

AIDS	acquired immunodeficiency syndrome
CAPI	computer-assisted personal interviewing
CI	confidence interval
DHS	demographic and health survey
DHS-7	the seventh phase of The Demographic and Health Surveys Program
DOB	date of birth
DV	domestic violence
FCT	field check table
HIV	human immunodeficiency virus
IA	implementing agency
KDHS	Kenya Demographic and Health Survey
KI	key informant
KII	key informant interview
MICS	Multiple Indicator Cluster Surveys
MOH	Ministry of Health
NFHS	India National Family Health Survey
NSO	National Statistical Office
PSU	primary sampling unit
SADHS	South Africa Demographic and Health Survey
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development

1 BACKGROUND

1.1 Objective

The Demographic and Health Surveys (DHS) Program provides technical assistance for the implementation of nationally representative household surveys in low- and middle-income countries. These surveys provide data for a wide range of population, health, and nutrition indicators. In some countries, where demographic and health data are lacking, these data are a primary source of important measures, such as child malnutrition, child mortality, and total fertility rate.

Since the first phase of The DHS Program in 1984,[1] DHS surveys have increased in scope and complexity as questionnaires are lengthened and survey modules[2] are added (Bradley 2016; Short Fabic, Choi, and Bird 2012). As questions and modules have been added to these surveys, research from resulting data has increased, suggesting that the growth in surveys is fulfilling global health needs (Short Fabic, Choi, and Bird 2012). However, there is evidence that increases in DHS questionnaire length are associated with a decrease in data quality (Bradley 2016). Increase in questionnaire length can increase interview time, and ultimately may present more burden for both the interviewer and the respondent. Many DHS surveys have included modules or additional topic-specific sections only among subsamples, resulting in questionnaires of differing lengths within the same survey. Individual interviews already vary in length because respondents are asked fewer or more questions based on life circumstances – for example, if a respondent has children or is married, she will be asked considerably more questions than if she is unmarried or has no children. However, in surveys where subsamples have one of two lengths of questionnaires, the quality of data from each subsample may be different if, in fact, the length of the questionnaires affects data quality.

DHS data quality has been the subject of analysis for some time. Though indicators themselves, such as anthropometry z-scores (Allen et al. 2019; Assaf, Kothari, and Pullum 2015; Perumal et al. 2020), fertility (Pullum 2019; Schoumaker 2014), or mortality (Hill et al. 2007; Stanton, Abderrahim, and Hill 2000) estimates, are often examined for data quality, age and date of birth remain the focus of many data quality analyses (Lyons-Amos and Stones 2017; Pullum 2019, 2006; Pullum and Staveteig 2017; Rutstein et al. 1990). Age and date-of-birth information are an ideal starting point to diagnose symptoms of poor data quality because they are difficult to obtain and are also the building blocks of many important demographic and health indicators (Pullum and Staveteig 2017). To understand DHS data quality, others have studied circumstances in fieldwork and interviewer characteristics (Johnson et al. 2009; Pullum et al. 2018). Early in fieldwork duration while nascent interviewers are learning, or late in fieldwork when they are fatigued, quality may suffer (Johnson et al. 2009). Older and more educated interviewers have been linked to higher data quality, and there is evidence that previous DHS survey experience is associated with both better and worse quality outcomes. While interviewers with previous DHS survey experience have lower levels of nonresponse, they also had substantially worse rates of heaping on ages ending in 0 or 5 (Pullum et al. 2018). It seems intuitive that longer questionnaires would have different effects than shorter questionnaires

[1] The DHS Program has spanned eight phases: DHS-I (1984-90), DHS-II (1989-93), DHS-III (1992-98), DHS-IV (1997-2003), DHS-V (2003-08), DHS-VI (2008-13), DHS-7 (2013-18), and DHS-8 (2018-present).

[2] Modules are sets of optional standardized, topic-specific questions that countries may choose to add in addition to the standard model questionnaire.

on fieldwork, interviewer fatigue and performance, and survey implementation. Surveys that have two different lengths of questionnaires offer an opportunity to explore the extent to which questionnaires of different lengths may have these effects.

This report aims to understand the effect, if any, of questionnaire length on data quality. We focus on two types of data quality indicators from surveys that utilized both long and short questionnaires: indicators that may reflect efforts on the part of fieldworkers to reduce survey burden (i.e., their workload) and those concerning age and date of birth that are notoriously difficult to collect in household surveys (Pullum and Staveteig 2017). Additionally, we explore themes resulting from qualitative interviews with experts who worked on surveys with long and short questionnaires to understand how these surveys are implemented and data quality considerations, if any, along the survey process.

This paper will employ quantitative data analysis using available data from the 2016 South Africa DHS (SADHS), 2014 Kenya DHS (KDHS), and 2015-16 India National Family Health Survey (NFHS), and qualitative analysis of interviews with key informants (KIs) who worked on those surveys to explore the following questions:

1. Are there differences in data quality between long and short questionnaires used in the same survey?

2. How are long and short questionnaires implemented in a survey and are there other elements, advantages, or disadvantages to consider when administering long and short questionnaires in the same survey?

1.2 The Survey Process

The survey process – and the circumstances under which the decision may be taken to use long and short questionnaires – may be somewhat opaque to those not directly involved in implementing surveys; for that reason, and because the key informant interviews discussed specific survey phases, this section contains a generic overview, not particular to any given survey.

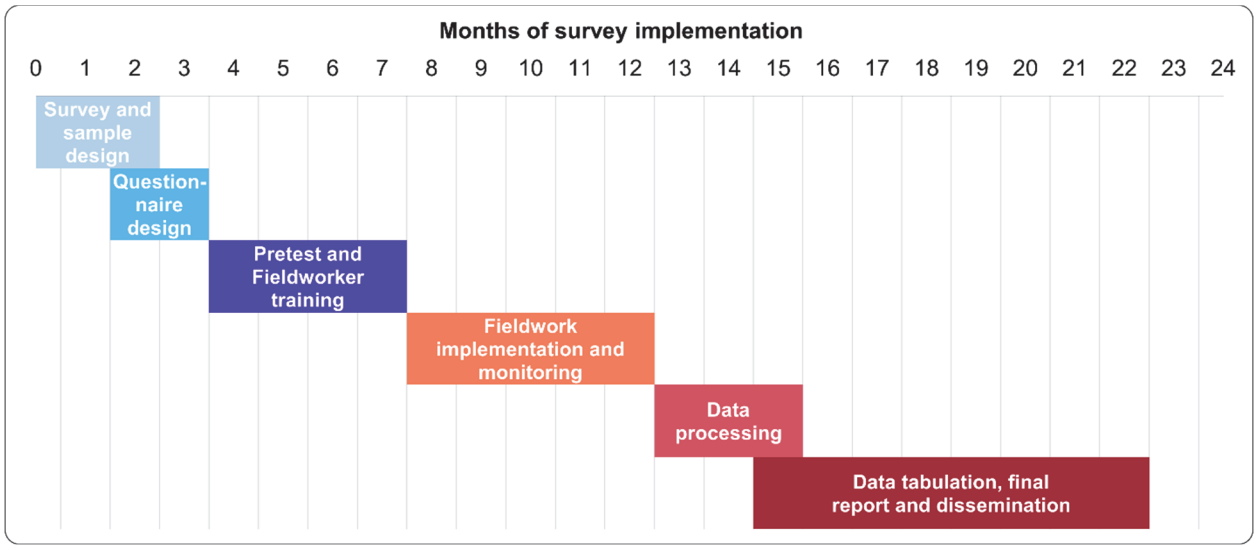
The overall responsibility for implementing a DHS survey generally resides with what is known as the survey implementing agency (IA), usually a country's National Statistical Office (NSO) or Ministry of Health (MOH). Staff of The DHS Program provide technical assistance to the IA at critical stages of the survey implementation to ensure that survey procedures are consistent with the DHS Program's technical standards and that the activities progress according to the timeline.

Many steps are required to ensure that survey data accurately reflect the situations they intend to describe, and that data are comparable across countries. The time to complete a survey depends on survey type, survey instruments, and sample size, but on average, DHS surveys take 22-24 months from design to results dissemination (Figure 1). Each DHS survey is executed in six phases:

Survey and Sample Design: The design for a DHS survey starts when a given country government (usually including the MOH and NSO and, often, the United States Agency for International Development (USAID) Mission in the country) concurs that it is time for a DHS. In a series of consultative meetings, stakeholders

(including The DHS Program) determine the overall sample size, geographic representation, general questionnaire content, biomarkers, timeline, and, crucially, donors and budget.

Figure 1 General DHS survey timeline in months



The objective of a standard DHS survey is to collect demographic and health indicators that are representative at the national level, for urban and rural areas, and usually for the first-level subnational administrative unit; these administrative units constitute the survey domains. DHS surveys use multistage stratified sampling. The number of sampling strata depends on the number of survey domains, with separate strata within each domain for the urban and rural areas. The target number of household interviews for a survey domain is based on complex demographic indicators, like fertility and mortality rates. These hard-to-estimate indicators generally require large samples to be estimated with precision. At this stage clusters are sampled, rather than households. A cluster is a geographic area that contains approximately 50-150 households; clusters are generally based on the country's census enumeration areas as the primary sampling unit. The DHS Program sampling expert reviews the sample frame or comprehensive list of clusters, designs the sample allocation of clusters per stratum, and selects the clusters within each stratum for fieldwork.

The budget for a DHS survey is also established at this phase and is subdivided by the major activities. Fieldwork is a major part of the overall cost, and the budgeting is based on six primary factors: 1) how many strata/sampling domains there are, 2) how many households must be visited in each stratum/sampling domain to achieve the desired number of interviews for representativeness for that stratum/sampling domain, 3) how many households will be selected per cluster (usually 20-35), 4) how many clusters are selected, 5) the number of fieldwork teams there will be, and 6) assumptions about how many days on average a team will spend in a cluster. Survey logistical needs, in addition to the large-scale and complex scope of each survey, require a large amount of funding. While a single donor may cover all the costs for a survey, increasingly multiple donors cover the survey costs, which requires a coalition of funders to work with the host country government to organize the required funding. The relationship between survey funding and survey design is iterative – specific amounts of funding are sought depending on the budget required to realize aspects of the survey, and aspects of the survey are revised depending on available funding. Putting together all the funding for a survey is a major undertaking before subsequent survey activities can begin.

Questionnaire Design: The second phase is deciding on the exact content of the survey questionnaires. Standard questionnaires (described below) form the basis for the questionnaires that are applied in each country. Adoption of the standard questionnaires allows all the major indicators to be produced on a comparable basis in different countries and across time within countries. The standard questionnaires are then adapted to reflect the reality and needs of the country. The DHS standard questionnaires are already lengthy and any additions need to be carefully weighed against increasing the overall length of the interviews. Generally, the longer the questionnaire, the more complex the training of interviewers will be and the longer and more taxing it will be on interviewers and respondents. Since many countries are interested in information in addition to what is included in the standard questionnaires, The DHS Program has developed several standard optional modules (described below). Use of these standard optional modules increases the comparability of data across and within countries, as well as saving time and effort in developing new questions.

Pretest and Fieldworker Training: The third phase involves training field staff on the survey procedures and questionnaire content. There are two components: a pretest and a main training. During the pretest, a smaller group of participants tests the questionnaires, field procedures, and computer-assisted personal interviewing (CAPI) system so that they will be as final as possible before the main training; the pretest also serves to review and finalize the training materials and agenda, and in some cases to train the IA staff who will conduct the main training. The main training takes place at least one month after the pretest to allow for the survey materials to be finalized prior to its start. The main training is a much larger exercise, during which those who will actually go to the field and collect data are trained; it also includes training for team supervisors. Both the pretest and main training include an element of field practice, a 'dry run' during which all the fieldwork procedures and questionnaires are practiced in clusters not selected for the actual survey. Before 2004, when CAPI was introduced in some surveys, all data collection was conducted on paper questionnaires. The use of CAPI has increased over time and surveys with fieldwork in 2015 or later are likely to use CAPI for interviews, although biomarker information is still collected on paper questionnaires and entered into the tablet computers at the end of each day.

Fieldwork Implementation and Monitoring: Fieldwork comprises the fourth phase of the survey. During fieldwork, eligible households and individual respondents are identified and interviewed. The term "fieldworkers" includes several roles. Team supervisors are responsible for managing team logistics and assigning work to members of their teams. Interviewers are responsible for conducting household and individual interviews. Depending on the survey and the country, there may need to be both female and male interviewers on a team; given the potential sensitivity of some questionnaire content, the usual approach is for female interviewers to conduct the Woman's Questionnaire and for male interviewers to conduct the Man's Questionnaire. The biomarkers collected for a survey are not always the same; biomarker technicians may measure respondent height and weight, conduct point-of-care tests (e.g., for hemoglobin, malaria antigens), and/or collect samples (dried blood spots, slides) for further laboratory testing (e.g., for HIV, malaria speciation).

DHS surveys generally employ several levels of fieldwork monitoring. The first level of supervision is provided by the team supervisors. The supervisors are responsible for closely monitoring the work of the teams to ensure that all selected households are visited and all eligible respondents are contacted. A second level of supervision consists of IA staff visits to the field. During these visits, interviews and data capture and editing are observed and problems or errors are discussed in review sessions with the teams. Finally, field check tables (FCTs) are produced periodically during fieldwork using data that have already been

captured. These tables look at issues such as response rates, the age distribution of household members, and measures of data quality for key questions. The tables contain results for each team, and any problems that appear from review of these tables are discussed with the appropriate team(s) and attempts are made to clarify issues, provide guidance, and ensure that problems do not persist.

Data Processing: The fifth phase involves rendering the data ready for tabulation and use. While some editing takes place simultaneously with data collection, there is still a process of editing and verification (e.g., consistency checking, analyzing and recoding 'other' answers, imputing missing information, calculating anthropometric indices, recoding variables, etc.) required to turn the fieldwork data ('raw') into standardized datasets ('recoded').

Data Tabulation, Final Report, and Dissemination: The final phase involves tabulating the data, preparing the final report, and disseminating the survey results in country. Final report tables are based on DHS standard tabulations, adapted to the specific survey questionnaires (for example, health facility types and levels are country-specific) and supplemented by country-specific tables for any nonstandard questions. Individual chapters of the final report are generally written by IA staff and other in-country stakeholders with support from DHS staff. The final report is usually presented at a national seminar where the main survey findings are shared with policy makers, program managers, researchers, nongovernmental organizations, and representatives of donor organizations. Complete survey datasets are usually made available to researchers on the same day as the national seminar.

1.3 Questionnaire Formats

The DHS Program maintains standard or model questionnaires that are adapted for each survey; these standard questionnaires are revised under each phase (5-year funding round) of The DHS Program; under the seventh phase (DHS-7), there were 5 standard questionnaires.[3] a Household Questionnaire, a Woman's Questionnaire, a Man's Questionnaire, a Biomarker Questionnaire, and a Fieldworker Questionnaire. In addition to the standard questionnaires, The DHS Program maintains standard modules that may be added to a given survey during the survey design/questionnaire design process. Under DHS-7, there were 12 such modules.[4] For an overview of the content of the standard questionnaires and modules, see Table 1.

The use of subsamples is a frequent practice in DHS surveys. Perhaps the most common example is the Man's Questionnaire, which is often only used in a subsample of households (perhaps half, or one-third). As discussed above, household and women's sample sizes are generally driven by the need to interview an adequate number of women per stratum to obtain estimates of fertility and child mortality; the indicators produced by interviewing men do not require such large sample sizes, so interviewing men in only a subsample of households is adequate for the data needs. The domestic violence module is also often deployed only in a subsample, although for different reasons: restricting it to women in the households where men are eligible for interview provides opportunity for further analysis that looks at data from both women and men, and restricting it to only one woman in those households is a practice guided by an ethical concern – reducing the risk of harm to a respondent by limiting the number of people in her household who know that this sensitive topic is being discussed.

[3] https://dhsprogram.com/publications/publication-dhsq7-dhs-questionnaires-and-manuals.cfm

[4] https://dhsprogram.com/publications/publication-dhsqmp-dhs-questionnaires-and-manuals.cfm

Table 1 Overview of DHS-7 standard questionnaire content and modules

Questionnaire	Standard content	Modules[1]
Household Questionnaire	Household Schedule: information on usual residents of the household and visitors who spent the night preceding the survey in the household, including age, sex, relationship to the head of household, education, parental survivorship and residence, and birth registration) Household characteristics: drinking water, toilet facilities, cooking fuel, household assets, and housing characteristics In countries with a high prevalence of malaria, the possession and use of mosquito nets	Disability (functioning in specific domains for all household residents/visitors age 5 and up) Road traffic accidents and other injuries Out-of-pocket health expenditures
Woman's Questionnaire *For women age 15-49 who are usual residents or visitors in the household.*	Section 1: Background characteristics (age, education, literacy, media exposure) Section 2: Reproductive history (dates and survival status of all live births, outcomes of other pregnancies in the past 5 years, current pregnancy status) Section 3: Contraceptive knowledge and use in the past 5 years Section 4: Antenatal care, delivery care, and postnatal care for births in the past 5 years Section 5: Child immunization. Section 6: Child health (recent diarrhea and fever) and nutrition (24-hour food recall) Section 7: Marriage and sexual activity Section 8: Fertility preferences Section 9: Husband's background, her employment status and ownership of assets Section 10: HIV knowledge/attitudes/behaviors and testing history Section 11: Other health issues (smoking, access to care, insurance)	Adult and maternal mortality (sibling history) Blood pressure Domestic violence Female genital cutting (includes component for Man's Questionnaire) Fistula Male child circumcision Maternal health care Newborn care Noncommunicable diseases (includes component for Man's Questionnaire)
Man's Questionnaire *For men age 15-49 (or 54 or 59, etc., depending on the country) who are usual residents or visitors in the household.*	Section 1: Background characteristics (age, education, literacy, media exposure) Section 2: Reproduction (number of children fathered) Section 3: Contraceptive knowledge Section 4: Marriage and sexual activity Section 5: Fertility preferences Section 6: Employment status and ownership of assets Section 7: HIV knowledge/attitudes/behaviors HIV and testing history Section 8: Other health issues (circumcision, smoking, insurance)	Female genital cutting (counterpart to Woman's Questionnaire module) Noncommunicable diseases (counterpart to Woman's Questionnaire module)
Biomarker Questionnaire *For children age 0-5, women age 15-49, and men age 15-49/54/59 who are usual residents or visitors in the household.*	Height/length, weight, hemoglobin level, and dried blood spot collection for laboratory HIV testing	
Fieldworker Questionnaire *For survey fieldworkers.*	Background information (education, languages, etc.)	

[1] In addition to the standard DHS-7 modules, the United Nations Children's Fund (UNICEF) Multiple Indicator Cluster Surveys (MICS) round 6[5] questions on Child Discipline and Child Labor were available as modules for the Household Questionnaire and the MICS6 questions on Early Childhood Development were available as a module for the Woman's Questionnaire.

[5] http://mics.unicef.org/tools?round=mics6#survey-design

2 DATA AND METHODS

2.1 Data

To conduct this analysis, surveys with long and short questionnaires needed to be identified. However, surveys that have subsamples with different questionnaires in each subsample often do not use terms such as "long" and "short" questionnaire, and there are no tracking mechanisms or markers to identify surveys that have employed such designs. Therefore, surveys were identified by asking The DHS Program experts to identify surveys that were designed with two versions of the same questionnaire – one version considerably longer than the other – to exclude surveys with minor differences between questionnaires (e.g., the inclusion of just one or two extra modules in a subsample, as is often done with the domestic violence module). Three surveys emerged from this process: the 2016 South Africa DHS, 2014 Kenya DHS, and 2015-16 India NFHS.[6]

As mentioned above, the standard questionnaires of The DHS Program are revised under each 5-year project phase, and surveys are often categorized by the phase to which they belong; this has implications not only for the questionnaires that they use, but for other details of data management and perhaps fieldworker roles and fieldwork procedures. The 2016 South Africa DHS is a DHS-7 survey; the 2014 Kenya DHS is a DHS-6 survey; and the 2015-16 India NFHS is a DHS-6 and DHS-7 hybrid. While each phase is different, there are enough consistencies between phases to allow the surveys to be considered together.

These surveys offer variety in size and scope to this analysis. The 2016 South Africa DHS had 8,514 households, the 2014 Kenya DHS survey sampled over 30,000 households, and the India 2015-16 NFHS had nearly 700,000 households (Table 2). South Africa and Kenya conducted the longer version of their questionnaire in half of the households in every cluster, while India conducted the longer questionnaire in half of the households in 30% of clusters.

Table 2 Summary of DHS surveys included in analysis

	2016 South Africa DHS	2014 Kenya DHS	2015-16 India NFHS
Total households Interviewed	11,083	36,430	601,509
Total women interviewed[1]	8,514	31,079	699,686
Percentage of clusters where long questionnaire was administered	100%	100%	30%
Percentage of all households interviewed with long questionnaire	50%	50%	15%
Women interviewed with long questionnaire[1]	4,193	14,741	122,351
Women interviewed with short questionnaire[1]	4,321	16,338	577,335

[1] Only women age 15-49 are reported here. In the 2016 South Africa DHS a larger age range of women was interviewed using only the long questionnaire, and those women are excluded in this analysis.

[6] The 2015-16 India NFHS terminology for the long and short questionnaires was state and district modules, respectively.

2.2 Methods

2.2.1 Quantitative methods

We focus on two types of data quality indicators: several that may reflect efforts on the part of fieldworkers to reduce survey burden (i.e., their workload), and those concerning information that is notoriously difficult to collect in household surveys. See Table 3 for indicators and their definitions.

Survey burden indicators

The first type of data quality indicator, survey burden indicators, are usually reflected in skipped portions of the questionnaire or shifting ages of household members out of the age range for eligible respondents. Some of these indicators can be monitored during fieldwork, while others are more appropriate for post-survey analysis.

Age displacement ratios. A standard DHS survey interviews women age 15-49, who are identified when a complete roster of household members is filled out in the Household Questionnaire. Interviewers may be tempted to reduce the number of women eligible for individual interview in a household in order to reduce their overall workload, and this temptation may be especially keen when women are on the borders of eligibility (age 15 or age 49). To see if women's ages have been shifted so that they are out of the eligible age range for interviews, a ratio is used to compare women who are on the boundary of the eligible age range to women who are one year of age outside the eligible range. The lower boundary age of eligibility is age 15, and the expected number of 14 and 15-year-olds should reasonably be the same (a ratio equal to one). If women are displaced out of the age range, from 15 to 14, the ratio of out-of-range to in-range would be larger than one. This logic can be used for the upper boundary of eligibility at age 49, and the expected number of 50- and 49-year-old women should be nearly the same. If women are displaced out of the age range, from 49 to 50, the ratio of out-of-range to in-range would be larger than one. It would be expected that the ratio of women age 14 to age 15, or the ratio of women age 50 to age 49, would not be substantially different among households that received the long questionnaire and the short questionnaire, and a difference may reveal where interviewers tried to displace eligible women.

Average number of eligible persons in a household. Every household is different and there is a range in the number of women eligible for the women's questionnaire in each household. Similarly, households will have a range in the number of eligible children under age 5 who will be weighed and measured during biomarker collection. The average number is used as an aggregate measure. It would be expected that the average number of eligible women or children should not differ among households that received the long questionnaire and households that receive the short questionnaire, and a difference may reveal where efforts were taken to reduce the number of eligible persons.

Skipping questionnaire sections. We created several indicators to examine if there is any evidence that entire sections, or long stretches, of the questionnaire were skipped. There are filter questions that determine whether an entire section or several questions of the questionnaire are skipped. In this analysis we chose three of these filters to serve as flags that questions or a questionnaire section may have been skipped: knowledge of HIV, use of a family planning method, and privacy for the domestic violence module. If a woman says that she has never heard of HIV, she is not asked any of the HIV/AIDS-related knowledge, attitudes, and behavior questions. Similarly, if a woman is not currently using any form of family planning

method, she is not asked a series of questions about the method(s), the duration of use, source(s), and whether she received information about side effects/other methods. Finally, the domestic violence (DV) module is skipped if there is not sufficient privacy to conduct this section of the questionnaire. While all of these responses are valid (women may not have heard of HIV, do not currently use family planning, or interviewers may not have had enough privacy to conduct the DV section of the questionnaire), they should not occur at significantly differing levels among those who were administered the long versus the short questionnaire.

Difficulty indicators

The second type of data quality indicator, difficulty indicators, are ones concerning age and date of birth that are notoriously difficult to collect accurately in household surveys.

Completeness of date of birth (DOB) for live births in the 5 years preceding the survey. In countries where vital registration and literacy are far from universal, it can be legitimately difficult for women to recall the exact date of birth (day, month, and year) of all of their children; it can be particularly difficult to do so for children who are deceased (and women may have both emotional and cultural reasons for being reluctant to share details about deceased children). Fieldworkers are trained to help respondents with difficult-to-recall dates, such as date of birth. When a respondent has difficulty recalling a date, the fieldworker will attempt to help her establish its relationship to events with known dates. These events may include important local religious, political, or environmental events, births or deaths of other children, and marriage. If necessary, respondents can provide an estimate for this information but for some proportion of births, respondents may simply not know all this information. Interviewers are expected to make serious efforts aiding respondents to recall and estimate, at the very least, both the month and year of birth, especially for births that are more recent. There may be temptation, when interviewers know they are facing a longer interview, to conserve energy, spend less time on this time-consuming and arduous probing, and accept more missing data; but hopefully the proportion of births with complete month and year of birth should not differ by whether a woman participated in the long or short questionnaire. The dates of birth for all live births in the 5 years preceding the survey are checked for completeness of information. We restrict to births in the last 5 years because the month and year of birth are essential components to many health indicators for children currently under 5, such as nutritional assessments or vaccination coverage.

Completeness of woman's date of birth. Akin to date of birth for recent live births, it can be difficult for women to know the date of their own birth. When the date of birth indicator is incomplete, it is often because the woman reported her age and year of birth, but not the month of birth. Interviewers may feel the same temptation described above when using a long questionnaire, but ideally the proportion of women for whom the year and month of birth are complete should not differ by whether a woman participated in the long or short questionnaire.

Table 3 Data quality indicators

Indicators	Definition	Population	Questionnaire
Survey Burden Indicators			
Age displacement ratio 14:15	Ratio of female household members age 14 to female household members age 15	Female household members age 14-15 in households	Household Questionnaire
Age displacement ratio 50:49	Ratio of female household members age 50 to female household members age 49	Female household members age 49-50	Household Questionnaire
Average number of women eligible for interview	Average number of women age 15-49 eligible for an individual interview among households	All households	Household Questionnaire
Average number of children eligible for biomarker collection	Average number of children age 0-59 months old eligible for height and weight measurement among households	All households	Household Questionnaire
No knowledge of HIV	Percentage of women age 15-49 who report having never heard of HIV	Women age 15-49	Woman's Questionnaire
No privacy to ask about domestic violence	Percentage of women age 15-49 who were not asked domestic violence questions due to lack of privacy	Women age 15-49	Woman's Questionnaire
No current use of family planning method	Percentage of women age 15-49 reporting no current use of any family planning method	Women age 15-49	Woman's Questionnaire
Difficulty Indicators			
Complete DOB for live births	Percentage of all live births in the 5 years preceding the survey who have a month and year of birth reported	Births 0-59 months preceding the survey	Woman's Questionnaire
Complete DOB for women	Percentage of women age 15-49 months who have a month and year of birth reported	Women age 15-49	Woman's Questionnaire

Analysis

To understand how these surveys were implemented, we describe the length of fieldwork. To illustrate the size and undertaking of fieldwork, we describe the number of interviewing teams. To describe the length of fieldwork, we use the difference between the earliest date of woman's interview and the latest date of woman's interview. The number of fieldworkers is created by using unique interviewer IDs in the woman's recode file, and unique supervisor IDs in the woman's file is used as a proxy for number of teams. If any interviewer or supervisor ID had fewer than 10 cases, we assume these are data entry errors and do not count these IDs in the total. To describe the long and short questionnaires in each survey, the average number of non-missing variables per woman is used as a proxy for interview length. Additionally, we describe the length of interviews in minutes when this information is available. If women were interviewed for longer than 95 minutes, their interview is stored as 95 minutes long, so we can assume our description of the length of the interview is underestimated. Interviews that took more than one visit to complete have no length of interview recorded, and we show the percentage of all woman's interviews that required more than one visit.

The full Kenya sample was used for analyses, though sample restrictions were employed for the South Africa and India data. In South Africa, the longer questionnaire included an extensive adult health module,

which was also administered to women older than the standard 15-49 age range. For comparability, these older women were excluded from our analyses as none of them responded to the short questionnaire and cannot be compared with respondents who were given the short questionnaire. In India, the long questionnaire was only administered in 30% of the clusters. Only these clusters were used in our analysis, in order to have reasonable comparisons between short and long questionnaires.

Data were separated into a binary category for long and short questionnaires, and households, women, and children were categorized by whether the household was selected for a long or short questionnaire. Indicators were calculated for each survey by the long and short questionnaire distinction and bivariate statistical analyses were conducted using chi-square tests for testing proportions, Wald test for ratios, and t-test for means. A p-value cutoff of 0.05 was used to signify evidence of a relationship and 95% confidence intervals (CI) are shown.

Data quality estimates often do not need to be weighted because real estimates of what occurred during fieldwork (i.e., exactly what percentage of interviews did not meet a given data quality standard) can be more useful than weighted estimates that account for national representativeness. However, without weighting the data to account for nonresponse and the multistage survey design, the data lack accurate standard errors needed for significance testing. Therefore, all statistical tests and estimates presented in this analysis account for the complex survey design. All analyses were done using Stata MP 16.1.

2.2.2 Qualitative methods

Key informant interviews

Key informant interviews were conducted from July 21 to August 5, 2020. Informants were selected purposively based on their participation in the survey design, implementation, or fieldwork of the 2016 South Africa DHS, 2014 Kenya DHS, and/or 2015-16 India NFHS. More specifically, individuals who held one of the following roles in the survey were identified for each survey: survey designer, survey manager, survey sampler, implementing agency survey manager, or fieldworker. It is possible that some key informants held more than one role or participated in more than one survey. Informants were both internal and external to The DHS Program. Informants were contacted via email to request interviews and all interviews were conducted through phone calls or video calls. Of the 11 individuals who were contacted to request an interview, 10 agreed to participate, and one did not respond.

Interviews were conducted using a semi-structured guide specific to the key informant's role in the survey process. The interview guide (see Appendix 1) focused on key informants' experiences with and opinions of the survey design, implementation, and/or fieldwork aspects of the differing questionnaire formats. The key informant interview guide was developed by the study investigators and reviewed by two DHS Program staff members: one with experience in qualitative data collection, and the other with survey implementation experience. Interviews were audio-recorded and then transcribed. All interviews were conducted by Jehan Ahmed and Julia Fleuret.

Key informants gave verbal consent to the audio-recorded interviews prior to the beginning of the interview. Though interviews were transcribed verbatim and quotes were recorded and are presented in the results, no statements are attributed to any individual. All individuals who gave permission are listed in Appendix 2. The ICF Institutional Review Board approved this study through an exempt review.

Data management and analysis

Interviews were audio-recorded using Microsoft Teams. In one interview, technical difficulties led the second half to be recorded on Skype for Business. All interviews were transcribed using Temi, an online audio-to-text automatic transcription service. These transcriptions were reviewed for accuracy against the recordings by Courtney Allen and Jehan Ahmed. All recordings and transcriptions were stored on a SharePoint site that was only accessible to the study investigators. Before the interviews concluded, we developed a codebook consisting of data-driven codes. After the interviews were conducted and transcribed, the content of the interviews was coded by Courtney Allen using NVIVO 11.

3 QUANTITATIVE RESULTS

3.1 Descriptive Results of Quantitative Analyses

Each of the three surveys had questionnaires that consisted of differing sections, modules, and questions. An overview of data collected in long and short questionnaires can be seen in Table 4. While there are some differences in the household or biomarker questionnaires for all three surveys, the bulk of the differences lie in the long and short woman's questionnaires. The 2016 South Africa DHS had the smallest difference between the long and short questionnaire. The long woman's questionnaire had only one additional (substantial) module. In Kenya and India, the differences in length and content between the long and short woman's questionnaires are more substantial. In the 2014 Kenya DHS, the woman's long questionnaire had eight more sections or modules included than the short questionnaire, while the India NFHS woman's long questionnaire had a difference of four sections.

Table 4 Overview of data collected in long and short questionnaires

	2016 South Africa DHS		2014 Kenya DHS		2015-16 India NFHS	
Household Questionnaire	**Long**	**Short**	**Long**	**Short**	**Long[5]**	**Short[5]**
Composition (e.g., headship, size, age, sex, education)	*	*	*	*	*	*
Characteristics (e.g., source of water, type of sanitation facilities, exposure to second-hand smoke)	*	*	*	*	*	*
Wealth index	*	*	*	*	*	*
Household ownership and use of mosquito nets	*	*	*	*	*	*
Household ownership of dwelling, land	*	*	*		*	*
Anthropometry of women age 15-49[1]	*		*		*	*
Anthropometry of children age 0-59 months[1]	*		*	*	*	*
Disability[1]	*	*				
Woman's Questionnaire	**Long**	**Short**	**Long**	**Short**	**Long**	**Short**
Individual characteristics (e.g., age, sex, education, marital status, media exposure)	*	*	*	*	*	*
Reproductive history [fertility and childhood mortality]	*	*	*	*	*	*
Knowledge and use of family planning methods	*	*	*	*	*	*
Fertility preferences	*	*	*		*	*
Antenatal, delivery, and postnatal care	*	*	*	*	*	*
Breastfeeding	*	*	*		*	*
Vaccinations and childhood illnesses	*	*	*	*	*	*
Infant and child feeding practices	*	*	*		*	*
Marriage and sexual activity	*	*	*	*	*	
Woman's work and husband's background characteristics	*	*	*		*	
Awareness and behavior about HIV and other sexually transmitted infections	*	*	*	*	*	
Domestic violence[2]	*	*	*		*	
Female circumcision[2]			*			
Fistula[2]			*			
Adult and pregnancy-related mortality[2]	*	*	*			
Adult health[3] *(women 15+ and men 15+ eligible)*	*					

Continued…

Table 4—*Continued*

	2016 South Africa DHS		2014 Kenya DHS		2015-16 India NFHS	
Men's Questionnaire	**Long**	**Short**	**Long**	**Short**	**Long**	**Short**
Full men's interview[4]	*		*		*	

[1] Anthropometry and other biomarkers may have been recorded in the household questionnaire or may have been recorded in a separate questionnaire (biomarker questionnaire). It is grouped here with the household questionnaire for ease.
[2] Sections in *italics* signify modules.
[3] Tobacco, alcohol, and codeine-containing medications; consumption of fat, salt, sugar, fruit, and vegetables; health care-seeking behaviors; and self-reported prevalence of a variety of noncommunicable diseases. This is the substantial additional module for the 2016 South Africa DHS.
[4] The Man's Questionnaire often collects information similar to that contained in the Woman's Questionnaire but is shorter because it does not contain equivalently detailed questions on reproductive and contraceptive use history or questions on maternal and child health, immunization, or nutrition. Men's data were not explored in this study and therefore are not detailed here.
[5] The 2015-16 India NFHS terminology for the long and short questionnaires was state and district modules, respectively.

These surveys were different not only in scope, but in size. The South Africa DHS took approximately 22 weeks of fieldwork to collect data and had 30 fieldwork teams consisting of 90 interviewers who administered the woman's long and short questionnaires (Table 5). The survey in Kenya was larger, with 25 weeks of fieldwork and 49 fieldwork teams staffed by over 150 interviewers. Though only the 30% of clusters in the India NFHS where long and short questionnaires were administered are described in Table 5, even this reduced survey remains remarkably large. This survey was in the field for 98 weeks, and 345 interviewers on 94 interviewing teams administered the long and short woman's questionnaires.

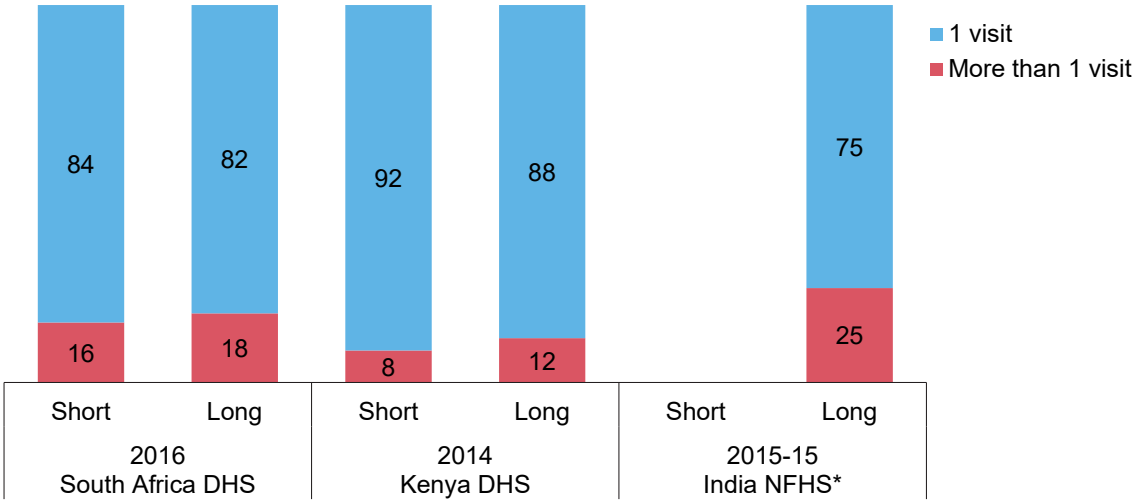
Table 5 Summary of teams, interviewers, and time spent in fieldwork in the 2016 SADHS, 2014 KDHS, 2015-16 NFHS

	2016 South Africa DHS	2014 Kenya DHS	2015-16 India NFHS[4,5]
Weeks of fieldwork[1]	22	28	98
Number of teams[2]	30	49	94
Number of interviewers[3]	91	158	345

[1] Weeks of fieldwork is the number of weeks from the first date of interview to the last date of interview.
[2] Unique supervisor IDs with at least 10 or more cases are used as a proxy for number of teams.
[3] Unique interviewer IDs with at least 10 or more cases are used as a proxy for number of interviewers.
[4] 2015-16 India NFHS sample only summarizes the 30% of clusters that included long and short questionnaires.
[5] 2015-16 India NFHS was conducted in two fieldwork phases: January-December 2015 (Phase 1) and January-December 2016 (Phase 2); different states took part in different phases of the fieldwork.

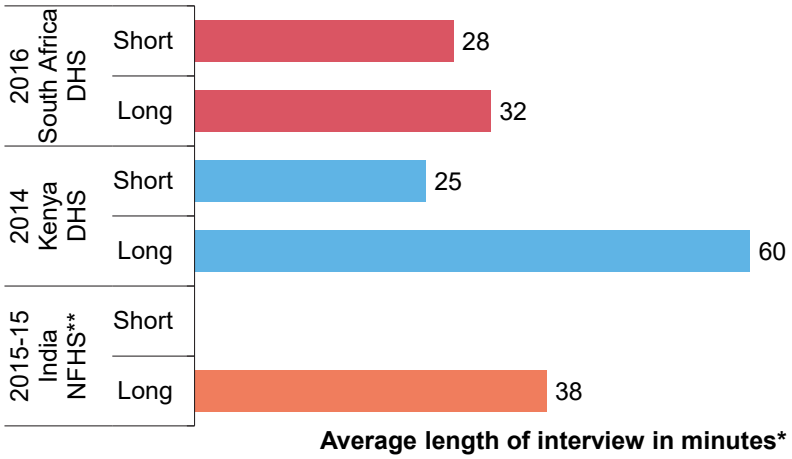
The distribution of woman's interviews completed in one sitting and the total length of time to complete the interview can be seen in Figures 2 and 3. The average length of a woman's interview in the South Africa DHS was 28 minutes for the short questionnaire and 32 minutes for the long questionnaire. In 16% of the short questionnaires and 18% of the long questionnaires, more than one visit was required to complete the interview and no length of interview is recorded. There is a sizable difference in woman's interview length between short and long questionnaires in the 2014 Kenya DHS, where the average short interview lasted 25 minutes – less than half as much time as the average long interview (1 hour). In Kenya, 8% of short interviews and 12% of long interviews occurred in more than one sitting. In the India NFHS, no length of interview time was available for the short questionnaires. The long questionnaires lasted 38 minutes on average, and one in four of these interviews occurred in more than one sitting.

Figure 2 Percent distribution of woman's interviews with short and long questionnaires completed in one or more visits in the 2016 SADHS, 2014 KDHS, 2015-16 NFHS



* 2015-16 India NFHS results only describe the 30% of clusters that included long and short questionnaires; no data were available on length of interview in the short questionnaires.

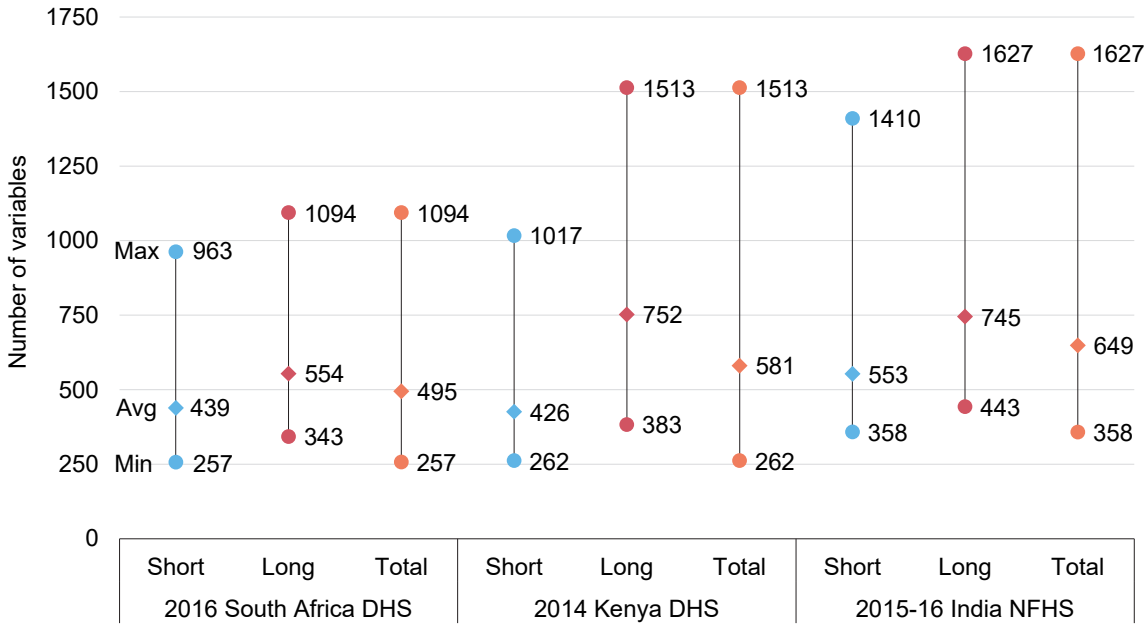
Figure 3 Summary of woman's interview length by short and long questionnaires



* Average length for interviews completed in one visit; if interviews lasted 95 minutes or longer, they are recorded as 95 minutes long.
** 2015-16 India NFHS results only describe the 30% of clusters that included long and short questionnaires; no data were collected on length of interview in the short questionnaires.

As a proxy for number of interview questions, the number of non-missing variables in the woman's individual recode was examined (Figure 4). As expected, the average number of variables in short questionnaires was always fewer than the average number of variables in the long questionnaire counterpart. In South Africa, the short questionnaire had an average of 115 variables fewer than the long questionnaire. The Kenya DHS had the largest difference in average variables between the long and short questionnaire, where the long questionnaire had an average of 752 variables – 326 variables more than the average short questionnaire. In India, the short questionnaire had an average of 553 variables – 192 variables fewer than the long questionnaire. The long questionnaires in the India survey had a maximum of over 1,600 variables.

Figure 4 Minimum, maximum, and average number of variables per woman in the 2016 SADHS, 2014 KDHS, 2015-16 NFHS



3.2 Data Quality Indicators

3.2.1 South Africa

In the 2016 South Africa DHS, there is no evidence that age displacement ratios are different between the long or short questionnaire (Table 6). While there is also no evidence of age displacement of women from age 15 to 14 or from age 49 to 50 in either the long or short questionnaires, as the confidence interval for these ratios includes one, the age displacement ratios are noticeably different between questionnaires. The mean number of eligible women age 15-49 to be individually interviewed was approximately 1.3 for both the long and short version of the questionnaire, with no evidence that the number of women differed by questionnaire type.

The proportion of women who reported not having heard of HIV was approximately 4% in each questionnaire type, with no evidence of interviewers possibly skipping the HIV section more frequently in one type of questionnaire or the other. Similarly, there is no evidence that the DV section was intentionally skipped more frequently in either the long or short questionnaire, with a similar proportion of interviews recording insufficient privacy to conduct this section of the interview (long: 17.4%, CI [14.8%,20.3%]; short: 15.1%, CI [12.8%,17.8%]). Approximately one in two women reported not currently using a family planning method in either type of questionnaire. Finally, the completeness of date of birth information was nearly universal in both types of questionnaire.

Table 6 Data quality indicators by long and short questionnaires in the 2016 South Africa DHS

	2016 South Africa DHS				
	Long QRE	[95% CI]	Short QRE	[95% CI]	p-value[2]
Age displacement ratio of women age 14:15	0.95	[0.72,1.18]	1.22	[0.92,1.52]	
Age displacement ratio of women age 50:49	1.22	[0.78,1.66]	0.95	[0.61,1.29]	
Mean number of eligible children for biomarker collection[1]	--	--	--	--	
Mean number of eligible women for interview	1.32	[1.27,1.38]	1.29	[1.24,1.35]	
No knowledge of HIV (%)	4.5	[3.6,5.6]	4.3	[3.4,5.5]	
No privacy for DV (%)	17.4	[14.8,20.3]	15.1	[12.8,17.8]	
No current family planning (%)	52.6	[50.4,54.8]	51.3	[49.1,53.5]	
Complete DOB for all live births in last 5 years, month and year (%)	99.6	[99.2,99.8]	99.8	[99.4,99.9]	
Complete DOB for women age 15-49, month and year (%)	99.9	[99.6,100.0]	99.9	[99.8,100.0]	

[1] Biomarker collection was not included in the short questionnaire and cannot be compared here.
[2] p-value indicates strength of association between short and long questionnaire estimates. *p <0.05; **p<0.01; ***p<0.001; left blank if not significant.

3.2.2 Kenya

In the 2014 Kenya DHS, there is evidence of age displacement of women from age 15 to 14. In the long questionnaire and short questionnaire, the ratio is greater than one, revealing a deficit in 15-year-olds. However, there is no evidence of a difference between the 14:15 age ratios in the long questionnaire and short questionnaire. There is evidence of upward age displacement from age 49 to 50 in the long questionnaire (Table 7), where there are 1.56 50-year-olds for every 49-year-old (CI: 1.17-1.95), however there is no evidence that this estimate differs from the estimate for short questionnaires. The number of children eligible to be weighed and measured was on average 0.8 per household in either questionnaire. The mean number of eligible women age 15-49 was approximately 1.2 for both the long and short version of the questionnaire with no evidence that the number of women differed by questionnaire type.

The proportion of women who reported not having heard of HIV was approximately 0.3% in each questionnaire type with no evidence of interviewers possibly skipping the HIV section more often in one type of questionnaire or the other. The percentage of women who reported not using a family planning method in either type of questionnaire is nearly equivalent (long: 57.0%, CI[55.8%,58.2%];% short: 57.8%, CI[56.6%,58.9%]. The completeness of date of birth for recent live births was nearly universal in both questionnaires, though the fraction of percentage where they differed was statistically significant. This is because nearly all of the cases where there was incomplete date-of-birth information occurred in the short questionnaire. However, this occurred so rarely (in less than 1% of cases) that this difference did not affect the overall estimate of complete date of birth by questionnaire. Women's complete date-of-birth information was approximately 87% in both long and short questionnaires and therefore did not differ by questionnaire type.

Table 7 Data quality indicators by long and short questionnaires in the 2014 Kenya DHS

	2014 Kenya DHS				
	Long QRE	[95% CI]	Short QRE	[95% CI]	p-value[2]
Age displacement ratio of women age 14:15	1.25	[1.10,1.39]	1.21	[1.08,1.35]	
Age displacement ratio of women age 50:49	1.56	[1.17,1.95]	1.23	[0.95,1.51]	
Mean number of eligible children for biomarker collection	0.83	[0.8,0.85]	0.80	[0.78,0.83]	
Mean number of eligible women for interview	1.16	[1.13,1.18]	1.17	[1.15,1.2]	
No knowledge of HIV (%)	0.3	[0.2,0.4]	0.3	[0.2,0.4]	
No privacy for DV (%)[1]	--	--	--	--	--
No current family planning (%)	57.0	[55.8,58.2]	57.8	[56.6,58.9]	
Complete DOB for live births in last 5 years, month and year (%)	100.0	[99.9,100.0]	99.7	[99.5,99.8]	***
Complete DOB for women age 15-49, month and year (%)	87.5	[86.6,88.3]	87.4	[86.6,88.2]	

[1] Domestic violence was not included in the short questionnaire and cannot be compared here.
[2] p-value indicates strength of association between short and long questionnaire estimates. *p <0.05; **p<0.01; ***p<0.001; left blank if not significant.

3.2.3 India

In the 2015-16 India NFHS, there is evidence of slight age displacement of women from age 15 to 14 in the long and short questionnaire (Table 8). While the ratios of 14- to 15-year-olds are just under 1.0, indicating slightly more 15-year-olds than 14-year-olds, there was no difference related to the type of questionnaire. There is evidence of upward age shifting of women from age 49 to 50, with a ratio of 1.44 in the long questionnaire and a higher ratio of 1.53 in the short questionnaire of 50 to 49-year-old women. However, there is no evidence that these estimates are statistically different between the two types of questionnaires. The number of eligible children to be weighed and measured was on average 0.6 per household in either questionnaire. There was a slight, but statistically significant, difference detected between the two types of questionnaires; this difference should be interpreted with caution given how small the magnitude of difference is. The mean number of eligible women age 15-49 to be individually interviewed was nearly the same in each questionnaire type, 1.45 among long questionnaire respondents and 1.46 among short questionnaire respondents. The percentage of women who reported not using a family planning method in either type of questionnaire is nearly the same (long: 58.5%, 95% CI[58.0%,59.0%]; short: 58.6%, 95% CI[58.2%,59.0%]. Finally, the completeness of date of birth for births was nearly universal, while completeness for women's date of birth was approximately 90% with either questionnaire, with no evidence of a relationship between type of questionnaire and completeness.

Table 8 Data quality indicators by long and short questionnaires in the 2015-16 India NFHS

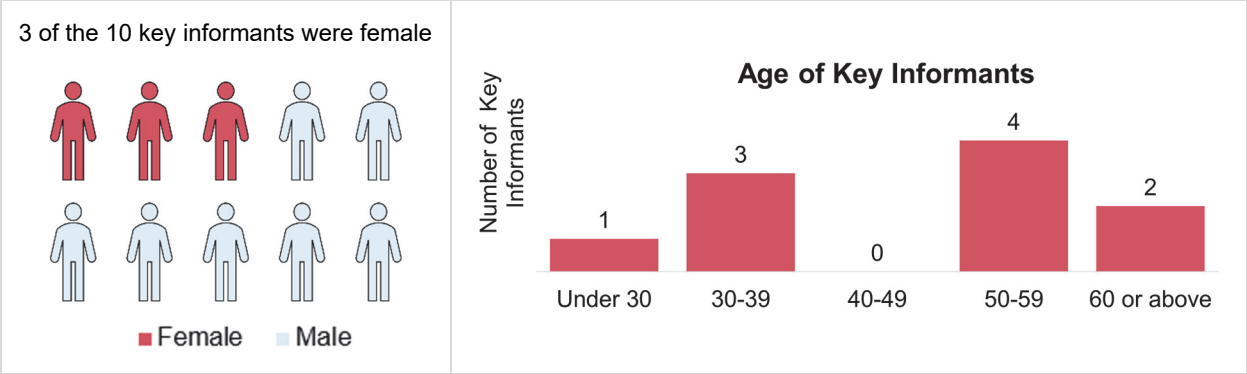
	India 2015-16 NFHS				
	Long QRE	[95% CI]	Short QRE	[95% CI]	p-value
Age displacement ratio of women age 14:15	0.92	[0.86,0.97]	0.94	[0.88,0.99]	
Age displacement ratio of women age 50:49	1.44	[1.30,1.58]	1.53	[1.40,1.67]	
Mean number of eligible children for biomarker collection	0.60	[0.59,0.61]	0.59	[0.58,0.60]	*
Mean number of eligible women for interview	1.45	[1.44,1.46]	1.46	[1.44,1.47]	
No knowledge of HIV (%)[1]	--	--	--	--	--
No privacy for DV (%)[1]	--	--	--	--	--
No current family planning (%)	58.5	[58.0,59.0]	58.6	[58.2,59.0]	
Complete DOB for live births in last 5 years, month and year (%)	99.6	[99.5,99.7]	99.6	[99.5,99.7]	
Complete DOB for women age 15-49, month and year (%)	89.8	[89.4,90.1]	89.7	[89.3,90.1]	

[1] The HIV section and DV module were not included in the short questionnaire and cannot be compared here.
[2] p-value indicates strength of association between short and long questionnaire estimates. *p <0.05; **p<0.01; ***p<0.001; left blank if not significant.

4 RESULTS FROM KEY INFORMANT INTERVIEWS

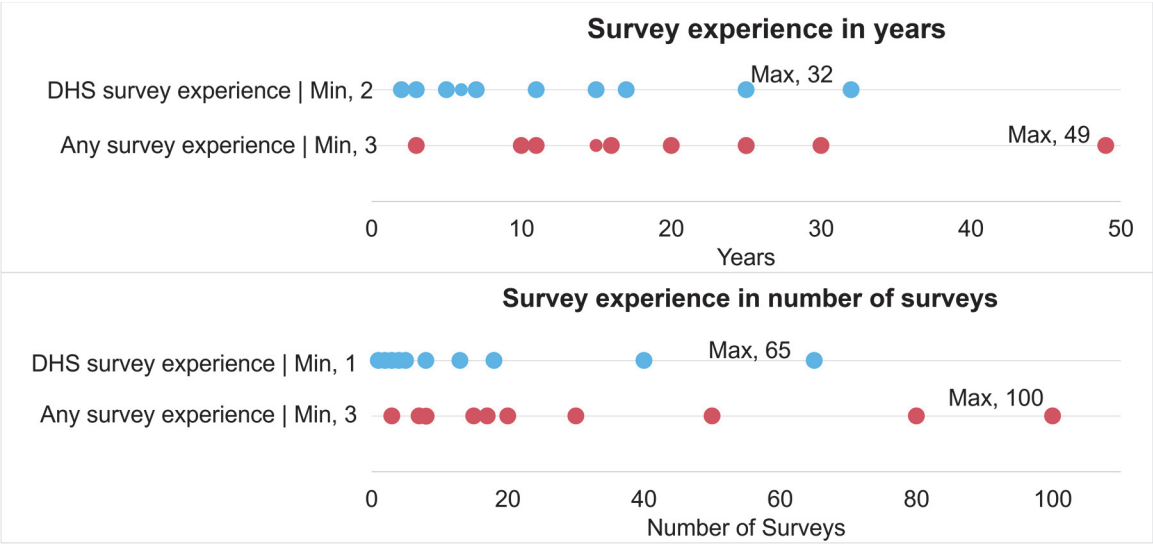
Key informants responded to an online survey that collected information on key demographic variables and survey experience. Three of the 10 key informants were female, and 60% were above the age of 50 (Figure 5).

Figure 5 Sex and age of key informants



The key informants represented a wide range of survey experience (Figure 6). While the minimum survey experience any key informant had was 3 years, the most experienced key informant had 49 years of experience working in surveys. When asked about the number of surveys the key informants had worked on, a wide range of experience was reported. Though some key informants reported having worked on 1 DHS survey, others reported having worked on 65 DHS surveys and 80 surveys overall.

Figure 6 Survey experience of key informants, measured in years and number of surveys



The key informant interview guide covered a number of survey phases (survey design, questionnaire design, sampling design, fieldworker training, fieldwork implementation, and fieldwork monitoring) and a few overarching topics (the informant's assumptions about and opinion of long and short questionnaires). The key informant interview guide was tailored to the informant's role on the survey, as not every survey staff

role is involved in every phase of a survey. Ultimately, the survey phase with which most respondents were involved for the survey(s) in question was fieldworker training (8), followed closely by fieldwork monitoring (7), sampling design (7), survey design (6), and questionnaire design (6); fewer informants had been involved in fieldwork implementation (2) (Table 9). For a more detailed breakdown of how many informants were asked each question, please see Appendix 1. Some ad hoc or improvisational questions were asked during interviews and are not included in this tally.

Table 9 Key informants' involvement in survey stages discussed during the interviews

	Number of informants involved in the survey phase			
Survey phase	2016 South Africa DHS	2014 Kenya DHS	2015-16 India NFHS	Total
Survey design	1	2	3	6
Questionnaire design	1	2	3	6
Sampling design	2	2	3	7
Fieldworker training	1	3	4	8
Fieldwork implementation	0	1	1	2
Fieldwork monitoring	1	2	4	7
Total informants for the survey	2	4	5	10

4.1 Key Informant Interview Themes

The key informants who were included in this study were asked about their experiences and opinions regarding the implementation of the long and short questionnaire formats in the surveys they worked on. While the interview guide was structured to discuss survey implementation phases in chronological order, informants were encouraged to elaborate on topics as they came up, allowing the conversation to unroll organically. The findings of these interviews are presented by theme, which were identified during the coding process. These themes are summarized with illustrative quotes; the quotes have been edited to remove fillers (e.g., 'uh') and identifiable details have been changed to more general terms while preserving meaning. Some of the themes uncovered during coding are closely related; some ideas expressed by respondents support more than one theme, and therefore some illustrative quotes are repeated in different (related) themes. The themes are summarized here in Table 10.

Table 10 Major qualitative themes of the key informant interviews

Qualitative Themes	
The case for having two questionnaires	
	Having a long and short questionnaire makes it possible to meet the need expressed by survey stakeholders for specific data while keeping survey cost within the available funding. It can also provide an opportunity to improve on past choices.
Decisions about survey and questionnaire design	
	Decisions about the sampling for long and short questionnaires are straightforward, but decisions about what content to include in which questionnaire can be complicated, and there will be extra considerations when time comes to finalize the data.
Survey burden	
	Having a long and short questionnaire results in a survey that is easier to manage; both fieldworker training and fieldwork monitoring are largely unaffected. However, some consideration is required during the day-to-day implementation of fieldwork (which questionnaires are tackled first, which questionnaires are assigned to which interviewers), and long questionnaires are more fatiguing to interviewers and respondents.
Explaining the concept of long and short questionnaires	
	There is no set definition for what constitutes a long and short questionnaire. If survey stakeholders do not fully understand the effect of the different sampling in long and short questionnaire surveys on indicator calculations, they may be left unsatisfied with the survey results.
Data quality	
	Data quality and its relationship with survey processes is an important consideration throughout the design and implementation of surveys with long and short questionnaires. There exists a conventional wisdom of an inverse relationship between questionnaire length, among other survey elements, and data quality, and certain data quality indicators can be monitored to identify issues during fieldwork.
Opinion and Future	
	While some key informants had reservations about having long and short questionnaires during survey design, once the implementation began, they observed their value and their usefulness for future surveys, either in setting a precedent or to find solutions for future survey design.

The results from the key informant interviews are presented by theme. A codebook was developed based on the key informant interview guide, which was refined during the course of coding; the resulting themes emerged from the coding.

4.1.1 The case for having two questionnaires

Cost

Cost can be a major stumbling block for survey planning.

"… our budget is extremely tight. [Funder] gave us a fraction of what they gave us in the previous survey, and sort of wanted us to do everything that we did in the previous survey. And… we said we can't. And, you know, we ask them for more money all the time and [stakeholder 1] did, and [stakeholder 2] did, and they just didn't have it." KII 1

There is a practical relationship between survey cost and the impetus towards using long and short questionnaires. Key informants frequently mentioned the number of questions, the time required to complete interviews, and ultimately therefore the duration of fieldwork, as a budgetary concern. Having a proportion of interviews that have fewer questions, and therefore should take less time to complete, should reduce the time needed to complete a cluster and the overall time (and money) required for fieldwork.

"I would say that the length of those sections was also a consideration… if you add all that up, there in the [long questionnaire sections], there were 209 questions of which 32 were filter questions. So, if you take out the filters, which are automatically done in CAPI, there were 177 questions, and it's a substantial amount of savings in time and cost of the field work." KII 1

"So we get bids… in terms of how many interviews, how many eligible households there will be… and they take the length into account when they're putting together their budgets… they generally plan, we say, they should spend three days a week in a PSU and if it's selected for the long questionnaire… we say they should spend five days in a PSU for each team." KII 1

"And that is when we first began redesigning the survey to accommodate the larger sample and among the data quality pieces or logistical or funding challenges that we mentioned. [Individual] proposed doing a long and short questionnaire in order to lessen the burden on the interviewers, because now, the survey instead of maybe the usual two to three months, it meant, it's going to about six months to cover these additional households… the workload, it's a lot, so it requires either more people or more time." KII 2

"Yes, the duration of the fieldwork was certainly an issue… I feel like we attempted to estimate how much the length of field work would change if we had a long versus a short. And that was somewhat hard to do. Like we weren't sure if it does it save us weeks, does it save us a whole month? Does it… decrease field teams' amount of time in a cluster by one day or so? So we had a really hard time getting an accurate estimate of how much time it would save us, but we felt that it would save us something…. And we also, I mean the logistics, it's also hard to find funding, to keep people in the field for a super, super long [time]." KII 2

"We always think once you have a larger sample size, you need to mobilize more resources." KII 6

"And also of course, now each questionnaire… each question, as you know, an additional question is a cost. So, it actually meant that, the traditional DHS will not be done because now… the cost was more than two or three times." KII 7

"So, but then to get the [lower administrative level unit] estimates, the sample size had increased six, four [times]… So, then it would be again [an] issue of time, money, resources, and everything." KII 9

"...because all the indicators… may not be relevant to have it at a [lower administrative unit] level, and if we needed them, then it would mean like we had to have an expanded sample size. So, it would definitely have saved on time resources and everything." KII 9

Balancing demands

There is also a political relationship between survey cost and the impetus towards using long and short questionnaires. There is a candid expectation that major data users and/or entities contributing towards the cost of the survey will be able to add questions or make other important decisions about the survey scope.

> *"So, basically [stakeholder] paid all the field costs… which is no small sum. And the problem with that is… it's their money, it's their survey. They're going to tell everybody what to do. Nobody can say boo about it." KII 1*
>
> *"And then another burden was that [funder] wanted to continue this massive [nonstandard set of questions]… So it was going to be a very, very long questionnaire, to begin with." KII 2*
>
> *"And then I think beyond that, yes, there was conversation about what can we do to… [funder] was adamant that they have measures at the [lower administrative unit] level." KII 2*
>
> *"I think [the decision to have long and short questionnaires] came for two reasons. One of which is that part of the leadership of the survey was [institution] under the leadership of [individual] and [that individual's] interests are in [specific subject matter]." KII 3*
>
> *"I mean, there were other factors that went into thinking how this module could be useful, but… come hell or high water, they were going to want to include these questions." KII 3*
>
> *"But we went to see [government official], [the official] said 'No, we need a [lower administrative unit] level survey.' I said, well, this already, we discussed. That we think based on the… implementing capability it's better to do a [upper administrative unit] level survey.... [The official said] 'I am not a statistician. I'm a politician. I must answer the request of the government.'" KII 6*

Deploying long and short questionnaires solves practical funding questions while satisfying political demands. Including long and short questionnaires in a survey enables a compromise between cost and providing what stakeholders want.

> *"So anything you can do to keep… interview length down on average, I think is a plus. But still getting the critical information that everyone needs or, thinks they need." KII 1*
>
> *"So a lot of times outside voices, other funders, other public health players at the table, want things added to a survey for their own use or their own programmatic needs that may… be valid interests, but really do add to the burden of a survey... If you do a long and a short version, then you can still receive the funding that they bring, meet their data needs, but lessen the burden on your field staff and on other quality and logistical concerns. So I think it allows more flexibility while also maybe being able to still fund a survey through these donors, who have demands." KII 2*
>
> *"… the idea of having short and long questionnaires or [having] modules for subsamples in general, sounds like [an] efficient solution, to basically, collect the necessary data only…. [So, it] has implications on the cost…" KII 5*
>
> *"I feel that [having long and short questionnaires] is very much useful as you can say, it reduced the cost because you can't ask all the people all the time… So compared to other survey, we are, I feel that we are in pretty good place to, estimate... So I feel that this is a very smart thing… I feel [the long and short questionnaires had a] very good impact in our survey." KII 10*

Aside from the funding aspect, it also allows for the satisfaction of unmet need. A DHS survey is often seen as a not-to-be-missed opportunity to obtain definitive data – on a specific topic, from a specific population, or for a specific geographic level. In addition to the standard topics covered by a DHS, there is often keen interest in adding more topics and more questions to produce more indicators. While a clear reason for having long and short questionnaires is to obtain lower administrative level estimates, other data needs are often related to special interests of stakeholders and contribute to the challenge of an ever-expanding questionnaire.

"I would say it's a way to allow the countries to, well, get whatever their priorities are onto the questionnaires, plus the standard DHS stuff, which we hope everybody asks everywhere." KII 1

"I think the Ministry of Health [was] driving [implementing agency] and others at the table to produce [lower administrative unit] level estimates because for them, estimates at that level were important for determining health budgets or determining the division or an allocation of resources." KII 2

"They want it because it's of interest to them. They're passionate about the topic." KII 3

"In that meeting various government ministry people used to come with a long list of questions, and then everyone wants to add the number of questions, no one is ready to delete some of the questions." KII 4

"This is an interesting solution to the challenge of everyone wanting all the data." KII 5

"So actually, if there would be no adequate [lower administrative unit]-specific indicators. Then the Ministry was saying where, the government was saying there was no need for the survey to be done." KII 6

"Why can't we do it? Why can't DHS… shorten the questionnaire? But afterwards I realized that it is not DHS. It is [the] Ministry people are very particular of every question they want to identify and tell that 'We are doing good' from our survey." KII 10

Sometimes, decisions about long and short questionnaires are an opportunity to improve on past choices. And it can be an opportunity to improve on past surveys and ensure that all the appropriate data are collected.

"In the past… they had done a survey many years before this one…. I think they had had a separate survey on like a separate questionnaire on [topic of interest]. That was a standalone questionnaire. So if a respondent was eligible for that questionnaire, they would do a [standard DHS questionnaire] for example, and the standalone questionnaire. However in thinking through how to do it this way, this time we proposed that they fold in… these additional topics into a module that could, through skip patterns, enable us to have it be seamless. Plus this was on CAPI. And what that would do that they didn't have last time was it would enable us to collect some of the key background information about the respondents…." KII 3

"I think there's one thing you need to pay attention is how to use the information to make the information more useful. For example, one example is the domestic violence [module]... But in the beginning, we try to put the domestic violence module in the half sample, where there is a no male survey. We try to balance the workload. But then we have questions, from the data users. They try to understand that the domestic violence with the characteristics of the partners. But we do not have the information, the data. So then we realized, Ah it's better we put the domestic violence in the subsample with the male survey, you collect the same amount of information, but with the male information, the information you collected is more useful. You can do really deep analysis… the design of the long and the short, really should put how to use the information collected into account." KII 6

4.1.2 Decisions about survey and questionnaire design

Besides the funding issue, there is also a technical case for not collecting extraneous data, and the decision making behind the sampling and representation of the long and short questionnaires can be fairly straightforward. Key informants agreed that there is no point in exceeding the required sample size for a given indicator; increased returns on the cost will be marginal. Specific indicators drive overall sample size, but many other indicators do not need the same sample size for precise estimates and surveys do not have to include them in every interview. If the sample size required for an indicator can be satisfied in a subsample, it is not efficient to include those questions in every interview.

"So the basic idea for sampling [is], to make sure you have enough cases, so that you can give the results at the national and [upper administrative unit] level or the national, [upper administrative unit] and [lower administrative unit] level." KII 1

"The bigger priority about what to include in either questionnaire was driven by statistics and what sample size you would need in the denominator in order to have some confident estimates. We need to have a certain number of pregnant women or a certain number of children under the age of five for some of our key measures… I want to say that those were the ones that were first priority [for] our considerations of what should be included. Will we have the ability to get key estimates, or will we have the ability to get estimates of key indicators at the [lower administrative unit] levels? And that's I think what drove the decision to include in both questionnaires… for some indicators... I mean, like maternal mortality is probably an obvious one. Like you're never going to get [lower administrative unit] estimates for that. So that wasn't even part of the discussion." KII 2

"It's a huge survey and it is a [lower administrative unit] level survey, but their aim for [specific topic] indicators wasn't actually to produce it on the [lower administrative unit] level or they want it, I believe for specific [upper administrative units]. So not all [upper administrative units] are affected the same way. So basically, I mean, having or applying the long questionnaire for all [upper administrative units] doesn't make any sense now because it will be [collecting] data that's not really needed, so it made sense in that sense to have a shorter survey for some households or for a subsample on a longer survey in areas where you are interested in producing… these specific modules, indicators." KII 5

"… when it comes to sample size, we use… the smallest sample size that can… measure the indicator with a good precision. So even if you have more budget, this doesn't really mean that we will increase the sample size. So if you want to produce this A indicator for X domain, we would recommend that you collect a sample size that exactly will give you that, that will give you the indicator with a good precision." KII 5

"… once actually you reach a specific sample size, the increase in precision after that, it's not like linear. So the increase in precision is minimal. So you might double the sample size and the increase in precision is like 2% – it's not efficient." KII 5

"…it's more about what you would like to measure. What are the key indicators, from these two surveys, from these two questionnaires and what is your target population, or what is the denominator of your indicator? And so, these are the two main questions that guide our decisions when it comes to the distribution, or I'll call it the allocation of the sampling units, over the sampling strata." KII 5

"So once you told me that we want to produce results on the national or regional [level], it's very clear in my mind what indicators you are talking about and what sample size I should consider." KII 5

"… [the sample design is] based on the common indicators, we can say main common indicators. Actually, our main concern is the child mortality and fertility rate for the sample size determination." KII 6

"Actually, for some indicators, you do not need a large sample size, so you can select a subsample." KII 6

"So, the Ministry of Health was one of the key stakeholders and we tasked them to tell us what would be the bare minimum, the key indicators that they wanted… at [lower administrative unit] level for us to include in the short questionnaires. So, we had… their representation at the steering committee level and also at the technical working committee level. So, from that understanding and the health policies, the documents on health policies, they get us a number of indicators…." KII 7

This is not to say that including long and short questionnaires solves all questionnaire design problems. After the decision is taken to include a given population or level of representativeness in a survey by having long and short questionnaires, specific decisions about which questions will be included in which questionnaire (and thereby which indicators produced for which populations/administrative levels), or the inclusion of country-specific questions, may still need to be taken. There is rarely much room to maneuver regarding concerns about the effects of those decisions on the survey, and the conversations can be contentious because stakeholders care greatly about data and data quality.

"I don't think [the stakeholder's questions] were added in the end and they were really, really angry about it… Oh, in the modules I remember fighting not to have [specific DHS-7 modules]. Those are not things that I would have ever advocated to include at the [lower administrative unit] level." KII 2

"… we put in [standard DHS-7 module], cause it does fit in, but we also had some perhaps ill-conceived questions on [topic of interest to in-country stakeholders]. So those were like questions that ordinarily we would have concerns about." KII 3

"So I know we kind of gave advice, and some of the questions we thought were not easy to tie to an indicator and sometimes we won. But not always. So, I have to say there's a lot of things in the questionnaire that… we weren't sure how well would work… we improve, tweak things to make it as hopefully as good as possible, but we had doubts about the validity of it; I think there was some concern, but that didn't stop it from being included." KII 3

"The questionnaire was changing up until the last possible minute and then in big ways… for example, [individual] felt very strongly we should not collect information on [county-specific questions]… [individual] felt like this was a huge waste of time, and it would be very hard to recode and all that. And they had done it before and [individual] thought it was not great, but there are a lot of big personalities here and… we didn't win that battle." KII 3

"But there are many, many incidences where we have said, no, that is not possible in this survey. On the basis of quality that if we are going to collect these data, that will not [be] good for quality… for example even [topic that was eventually included] questions, we didn't want, that has been imposed on us." KII 4

On occasion, a lack of interest in or negative opinions about certain questions or indicators may drive decision making.

"There were people all along, and probably not just in the [preceding survey], but even before that, that didn't want any sexual relations questions in the survey. They thought it was inappropriate. And, that probably helped to lead into that decision. If you're asking people a lot of sensitive questions, it's best to ask fewer people those sensitive questions and still get the information that you want." KII 1

"And so the sections that were taken out – there were four sections that were not included in the short questionnaire. And they tended to be either things that were not of primary importance, there were no standard international indicators or SDG indicators, and they… tended to be sections, which were sensitive in nature." KII 1

"Because the issues we are covering in long questionnaire, those issues are not required to be estimated at a [lower administrative unit] level. People are satisfied with [upper administrative unit] level estimates of those issues." KII 4

The decision to have long and short questionnaires, and then the decision about what to include in which questionnaire, is not the end of the line. Even after all the decisions have been made, and no matter how clear-cut some of them may have been, there may be challenges in handling the data including recoding country-specific questions, producing weights for nonstandard populations, and estimating sampling errors for nonstandard indicators.

"And thinking about data... I know there's issues with recoding this [country-specific] data… we don't have the structure as I understood it, what do we do with these nonstandard variables for nonstandard population? It was, I think, complicated… it's more from the tabulation standpoint and making data available, but there were complications there." KII 3

"What I remember in [country] was that they wanted that also to include the women beyond the age limit, older than 49 years old, which had some implications on the survey weights and how we produce or where we should keep the data of these women in the micro data files." KII 5

"I remember it was challenging… If there is country-specific indicators. So now we are talking about new indicators that I have to calculate, and estimate, the standard error and confidence intervals for." KII 5

4.1.3 Survey burden

A more complex design translates to a simpler survey to manage and implement, especially with CAPI. A survey with long and short questionnaires may initially be considered a more complex design than a survey with, generally, the same set of questionnaires administered in all households. However, key informants discussed that a larger survey requires more management and implementation support. For example, survey managers and implementers must work with larger budgets, more staff, and longer and larger scale fieldwork. In addition to being easier to manage and implement, with CAPI, training on these more complex designs is no longer an issue. On paper, the more complex survey design needed reinforcement during fieldworker training. With CAPI, the user experience is that these designs are easier to train on since the assignment of the interview for long or short questionnaire is automatic.

"The money actually to me, even if you have… the entire survey being at the long questionnaire, as long as they have the right management, we still get… the data quality. But, my worry would be how big should the management team for that [have to be] to be able to give… the required quality data from the [survey]?" KII 8

"I would say so. The idea of having short and long questionnaires or having modules for subsamples in general sounds like an efficient solution. Collect the necessary data only. It has implications on the cost and on the burden in general… I mean, the challenge before with paper [questionnaires] was how to train interviewers on dealing with these, complex designs. I believe, from my humble experience in the field, with CAPI, I believe that should be easier." KII 5

"You do mention at the beginning [of the training] that not everybody is going to answer all the questions. And then when you get to the last question in Section [X], which is the filter question, determining whether you go on, then they handle it like any other filter question. And of course in CAPI, it is seamless." KII 1

Fieldwork duration is likely shorter. One key informant noted that shorter questionnaires reduce a fieldwork team's time within a cluster or community, which in turn reduces the overall time for which they are conducting fieldwork. The greater the time in the field, the greater the health and safety risks are for the fieldworkers.

> *"Yes, the duration of the fieldwork was certainly an issue… I feel like we attempted to estimate how much the length of fieldwork would change if we had a long versus a short. And that was somewhat hard to do. Like we weren't sure… does it save us weeks? Does it save us a whole month? Does it… decrease field teams' amount of time in a cluster by one day or so? So we had a really hard time getting an accurate estimate of how much time it would save us, but we felt that it would save us something." KII 2*
>
> *"We did not want the field teams there indefinitely. We did not want them running into different seasons that would put them at greater risk for flooding or malaria or whatever else… And [for the logistics], it's also hard to find funding to keep people in the field for a super long [time]. There were definitely some violence or unsafe situations in which they were in. So we were cognizant of keeping teams in the field any longer than necessary." KII 2*

Because the short questionnaire is a subset, training focuses on the long questionnaire. Fieldworker trainings in surveys with long and short questionnaires are generally not a challenge, according to key informants. All trainees are introduced to the subsampling in the survey, but frequently spend the majority of the training learning about the content of the long questionnaire since the short questionnaires are often simply a subset of the long questionnaires. This training strategy is similar to what is frequently used for the woman's and man's questionnaires, since most of the man's questionnaire is identical to questions found in the woman's. Later in the training for some surveys, interviewers practice administering both long and short questionnaires, but in other surveys the practice is only on the long questionnaires. One area that has proven challenging is anxiety and doubt associated with the long questionnaire among both trainers and trainees – some are skeptical of the length of the interview being acceptable to the respondent, while others are nervous about administering such a lengthy interview. One key informant mentioned that trainees were told that both data quality and logistical challenges were considered as rationale for having both long and short questionnaires in the survey.

> *"[Having the long and short questionnaires] doesn't impact training greatly." KII 2*
>
> *"The training on the short and long questionnaire actually came later on. First, we went through the long questionnaire, because actually that one is just a subset of this one… After that, then we allude [that] actually we will not be administering it to all of the households, but just some of the households in the sample." KII 8*
>
> *"At the beginning of each of the trainings, I had a few slides that visually showed [what] previous DHSes had done. This DHS is huge, because it is so big we had to make some adjustments to help accommodate the data quality challenges and the logistical challenges of doing a survey… One of those is a long and a short questionnaire, and here's a visual of what the difference in each of them is. Here is a visual that will show you, as part of the field team, how you would determine where to administer a long and short questionnaire. [And], in the end, when we get to our reporting stage, here's the impact of a long and short questionnaire is going to have." KII 2*

Long and short questionnaires may be unequally distributed among interviewers. While CAPI is pre-programmed with the allocation of the long and short questionnaires for each household or individual, the team supervisor must decide how to distribute the interviews among their team members. In some instances, the long questionnaires may be assigned to stronger interviewers, but in others they may be assigned to

interviewers that the team supervisor does not get along with. Interviewers may compare the numbers of long and short questionnaires assigned to them and may harbor negative feelings if they feel the distribution has not been equitable. In one survey, weaker interviewers often started fieldwork by administering the short questionnaire, and gradually, under the mentorship and supervision of stronger interviewers on their team, became more comfortable administering the long questionnaires, too.

> *"Initially, there are a lot of problems during the [long and short questionnaire] distribution [between fieldworkers]." KII 10*
>
> *"I didn't want others [to] say I'm partisan, I'm only assigning short questionnaires to [some] enumerators while assigning long ones to others. I made sure that before we left the office, actually everybody had a mixture of both the short and the long questionnaire households to cover." KII 8*
>
> *"There was an internal partiality among the supervisor with the investigators… The time taken for the long questionnaire is more, so they were interested to give [them to] the person who is not friendly to them… The overall thing we were trying to summarize is that don't burden everybody, and don't count the number of [long and short questionnaires], to complete the task [as a] team." KII 10*

Longer questionnaires lead to interviewer, respondent, and community fatigue. All key informants discussed the effect of questionnaire length on fatigue, and, in turn, of fatigue on data quality. While interviewer and respondent fatigue were most commonly cited, fatigue experienced by the cluster or community where fieldwork is done was also mentioned.

Interviewer fatigue was frequently discussed as a reason for designing surveys with long and short questionnaires. Administering long questionnaires to every respondent seemed like an unnecessary burden on the interviewer, which would translate to a decline in data quality. While most key informants discussed interviewer fatigue as a singular phenomenon, their responses revealed two distinct forms of fatigue: with the questionnaires, and with the survey fieldwork as a whole. Fatigue with the questionnaires was highlighted in several KIIs. First, while CAPI improves linkages in information collected in different sections of the questionnaires, there is still some information the interviewer needs to remember from earlier in the questionnaire to properly administer later questions or sections. The longer the questionnaire, the less likely the interviewer is to remember the information collected earlier, thus affecting how later questions are asked to the respondents. Second, the longer the interview, the more likely it is to get disrupted, either because of the respondent's fatigue or because of an outside event (e.g., privacy being compromised, competing priorities requiring the respondent's attention, etc.). Interviewer fatigue with the survey fieldwork was discussed more generally, with key informants mentioning interviewer boredom with repeating the questions, or strain of being away from home for an extended period of time.

Respondent fatigue was also discussed as a reason for having the long and short questionnaires. Key informants mentioned the difficulty for interviewers of maintaining the respondent's attention for a long period of time. The respondent may start feeling bored or tired as the interview progresses and their fatigue may manifest in different ways. First, the respondent may start answering questions without fully considering the question or thinking about their response. Second, the respondent may wish to either postpone or end the interview. The risk of refusal thus becomes higher as the questionnaire length increases, which is a serious issue as considerable time may be invested in an interview only to have the respondent refuse and the data be discarded. A key informant said that some interviewers were confident of their ability to manage a certain amount of time with a respondent, but not more than that.

Community fatigue is the idea that a community's enthusiasm for the survey may decrease the longer the fieldwork team is in the cluster. Community acceptance for the survey may wane because of the collective buildup of individual respondent's fatigue and may in turn affect participation in the survey in the remaining days of the cluster work.

Relieving the burden is done, in part, by breaking up the interview and scheduling revisits. Interviewers and teams can share their strategies for dealing with burden with each other.

"Having a short version of the questionnaire would lessen field time, would lessen the burden on interviewers, to some degree lessen respondent fatigue, in that you spend maybe less time in a cluster or within the community or annoying the respondents. So not like respondent fatigue, [like] whoever received the short questionnaire ends up talking for 60 minutes and whoever receives the long questionnaire end up talking for 90 minutes, but just the broader sense of respondent fatigue." KII 2

"Long questionnaires will often take upwards of four hours to complete all the questionnaires, the burden on the respondents may discourage participation or incline respondents end interviews before the questionnaires have been completed." KII 3

"The reason why they told us they were bringing about the short and long questionnaire was because of the fatigue… by having the interviewers getting fatigued by administering the long questionnaire to very many households because we had a very large sample size for the DHS." KII 8

"We were wondering about [whether the] investigator will be able to cover all these things, keeping in mind sequence relating to the first question, like you want child, number of [children] and [whether] you want any more [children] in the next section, if they can relate and talk to them, keeping in mind how many children they have in that house." KII 10

"If you occupy [respondents for a] while, you know, for hours, they may get bored [towards] the end. They try to give you answers without… really carefully thinking. Theoretically, it's always true." KII 6

"By reaching Section 10, Section 11 after interviewing for over an hour, there is interviewer fatigue. Some may intentionally [choose] that [the respondent] has never heard of HIV/AIDS. Then they are skipping forty questions." KII 4

"I felt like after the long questionnaire, the short questionnaire is not much of a problem after listening to [the interviewers'] experiences… When the debriefing was going on, we used to talk about these types of things which they handled because [in some other team], they were not able to handle. They were saying that there was more non-response halfway… that we can't give you more time. So, refusal, so all data will be [gone]." KII 10

"Your respondents will get really tired and really bored. And in the communities that they're within, [they] may not necessarily continue to welcome you if you remain there at length." KII 2

Long and short questionnaires may be administered differently. In practice, the longer a questionnaire is, the more likely an interviewer is to encounter difficulties in administering it. There are several reasons interviewers encounter challenges with the long questionnaire. First, people can walk in and out of the interview space, thereby disturbing the privacy needed to administer most of the questionnaire. Second, in surveys where a larger proportion of respondents get the short questionnaire, interviewers are more comfortable administering the short questionnaire than the long questionnaire. In addition to there simply being less content in the short questionnaire, the interviewers have administered so many that it has become second nature. One key informant mentioned some interviewers not even needing to look at the short questionnaire when asking questions because they knew it well enough to administer with minimal error.

Then, for sections exclusive to the long questionnaire, interviewers had to go more slowly and read the questions. Third, in order to get through the interview quickly, the interviewer may not ask questions properly, not record answers appropriately, or, in extreme cases, may deliberately or very easily give up, not getting privacy or consent. One key informant mentioned that some interviewers would 'kill off' or deliberately not list an eligible woman in the household questionnaire to reduce the number of long woman's questionnaires they would need to administer. Finally, one key informant mentioned that for long questionnaires there would be less time between the completion of the interviews and the exit from the cluster. As a result, the long questionnaires would sometimes be rushed, or they would be completed right at the end of the cluster work, leaving little to no time for the team supervisor to resolve errors in the data through data quality checks.

"They found it easy when they administered the short questionnaire because they would take a very few minutes to complete the interview. Unlike when they are going through the long questionnaire… Because now there's a situation whereby the enumerators who would like to "kill" the women in the household, whenever there's a long questionnaire so that they don't have to administer the very many long questionnaire to the women in the household." KII 8

"I could see that the short one is very short and easy for me to administer." KII 8

"With regard to domestic violence – sometimes you would find that [interviewers] are deliberately or very easily giving up and not covering it." KII 9

"Yeah, look at [this FCT indicator]. Eligible women not interviewed due to lack of privacy. This is a clear example of interviewer fatigue… Longer version also has more tendency of skipping something." KII 4

"I don't think I noticed any [differences] necessarily in implementation or in energy level, except for that, when [interviewers] knew they had a long women's questionnaires, they would have really hunkered down… And when they had a short questionnaire, it was just like – I got this, don't worry." KII 2

"[Those interviewers who were less experienced] were uncomfortable. They were not… confident enough to ask those questions. Like I can manage that much time of the household respondent… If you go into some more time or the husband comes in between. So, any practical problems [that] will be coming… they were really very much afraid to face those situations." KII 10

"The fieldworkers were happy to have a short questionnaire They would say things like 'when I know it's a short household, I get really happy.' Or you know, 'these ones go really quick and the other ones take forever'." KII 2

Field check tables and monitoring efforts remain largely unaffected. Generally, fieldwork for the long and short questionnaires can be adequately monitored through the standard DHS FCTs. Standard indicators in the tables can be used as proxy to find out how many long and short questionnaires have been completed, and nonstandard ones can be added to provide more information on indicators of interest. Fieldwork monitors can, through FCTs as well as through in-person monitoring visits, determine whether the questionnaires are being equitably distributed or if certain interviewers are overburdened. Monitors sometimes have to be prescriptive with questionnaire assignments to prevent team supervisors from assigning based on favoritism. Finally, ensuring confidence and capacity of all interviewers in administering the long questionnaire is important, as is emphasizing the completion of a team-based task over the number of long and short questionnaires completed per interviewer.

"In field check tables, there is nothing like that - How many people have done [the long questionnaire]? How many teams have done [the long questionnaire]? How many teams have done [the short questionnaire]? For that I used to go to the male interviews. So male interviews are done only in the [the long questionnaire]… Okay. So how is, how is their response rate?" KII 10

"Within the team [if] supervisors are being partial to people, [sometimes we] interfered and said that, okay, send this person, send that person [to administer the long or short questionnaires]." KII 10

"[I used to tell the interviewers administering mostly short questionnaires:] Now it's time for you to do the [long questionnaire]. You have done enough [short questionnaires, are] confident enough. Now go to the next level… Once the investigator is confident and really happy to do it, then we don't need to worry in the later stage." KII 10

"I didn't want others [to]say I'm partisan, I'm only assigning short questionnaires to [some] enumerators while assigning long ones to others. I made sure that before we left the office, actually everybody had a mixture of both the short and the long questionnaire households to cover." KII 8

4.1.4 Explaining the concept of short- and long-form questionnaires can be difficult

There is no definition for what constitutes a long and short questionnaire. Key informants discussed the inadequacy of using words like 'long' and 'short' to describe the different questionnaires. In the Kenya DHS, the two versions of questionnaires were termed as full and short, and neither the India NFHS nor the South Africa DHS used 'long' and 'short' terminology. The India NFHS used different terminology, 'state' and 'district' modules, while the South Africa DHS used no official terminology as their longer questionnaire only consisted of one extensive additional module. Some key informants noted the lack of a clear definition of what is considered long and short – noting that many DHS surveys have modules added to the questionnaires in a select subsample.

"I just should state at the outset that almost nobody ever talks about this as a 'long' and 'short' questionnaire." KII 1

"So I want to say it was like a long and longer questionnaire, there was no real short [questionnaire]…" KII 3

"Actually, in many surveys you have short and long questionnaire." KII 6

Explaining to stakeholders. While the long and short questionnaire approach might offer a solution when balancing demands to include lower administrative level estimates (e.g., district estimates), key informants found that survey results were still met with demands when stakeholders realized that not all indicators could be calculated at the same level. Those who implement surveys are tasked with managing expectations of stakeholders when discussing what data will result from having two questionnaires. Understanding and explaining what is and is not possible when using long and short questionnaires was an added element of complexity that survey implementers had to navigate.

"…if there was some way for people to have better remembered or better manage their expectations for what could have been available after doing a long or short questionnaire." KII 2

"So if you are not very clear of that very beginning, then again also passing it to the others [is] difficult or making people understand what is the difference between this, and this, and how will it be actually be actualized or implemented." KII 7

Despite it all, sometimes at the end of survey implementation, not everyone is satisfied. Unfortunately, not all parties have a crystalline understanding of the relationship between a questionnaire, the sample in which it is implemented, and what can be tabulated. This can lead to disgruntlement when the survey report is being written, or when the data are released. Additionally, secondary users, to whom it may not be immediately apparent that different indicators were collected from different people, may have understandable challenges conducting accurate analysis.

"It was a little bit uncomfortable when we got to the final report and [stakeholders] had either forgotten or misremembered what was going to be at the [administrative unit] level and what would not be… more than one individual, not just [funder], but other individuals being upset that certain indicators were not available at the [lower administrative unit] level." KII 2

"… that representative wanted the [specific DHS-7 module] and incidentally were really upset when the findings from that were so minimal." KII 2

"And I can't think of any indicator that I thought was mistakenly categorized, mistakenly available. Like I was pretty satisfied, but there were definitely other stakeholders who would have liked to have seen X, Y, or Z or didn't understand statistically why it's not possible. So those conversations were not comfortable and I wish I could have avoided them... So if there was some way for people to have better remembered or better manage their expectations for what could have been available after doing a long or short questionnaire." KII 2

"I remembered going back and forth… at the report writing stage and being like, are we really sure that we want to present this? Or what footnotes do we need to add to this table so that people don't misinterpret this." KII 2

"Just to say that that was another issue with the [extra content] was that no thinking had been done ahead of time about, really about what, how this, how this should be tabulated... And so that's just another consideration when having long and short questionnaires is what the effects are going to be on the tabulation... But in this case, it was very frustrating that we got to report writing and [stakeholders] were unhappy with the tabulations or hadn't spent much time thinking about them." KII 3

"And today many, many results… [data users] ask us the question that why [specific topic] information has been blank in 85% of cases because they have not studied our design. They're not aware." KII 4

"… when people then try and analyze data… until they get really familiar and read all the [documentation] they don't realize that certain indicators are not for the [lower administrative unit] level. So, they analyze and make interpretations because the data has those numbers, but the Ns, like, you know, people who won't go into the [analysis] with clear understanding." KII 9

4.1.5 Data quality

Data quality is a significant preoccupation in the design and implementation of long and short questionnaires. Key informants discuss the role of data quality in every phase of implementation and often with specific mention of long and short questionnaires. Key informants mention that decisions about which questions are contained in the long or short questionnaires or how the questionnaire types are distributed among interviewers are made with data quality in mind.

"Whatever is required at [upper administrative unit] level. We put it in larger version of questionnaire, whatever is required at [lower administrative unit] level, we used to put at that. In that discussion also, many times quality of data becomes paramount, important." KII 4

"We, we are calling 'innovation in data quality'. Survey innovation in survey implementation, that we have changed during this survey. And probably 16 new items we have added." KII 4

"… when it comes to sample size, we use… the least sample size or the smallest sample size that can give you a good, that can measure the indicator with a good precision. So even if you have more budget, this doesn't really mean that we will increase the sample size. So if you want to produce this an A indicator for a X domain, we would recommend that you collect a sample size that exactly will give you that, that will give you the indicator with a good precision. And when we are doing this, we have the quality in mind. So regardless of the budget, we have the quality, data quality in mind." KII 5

"We certainly did not want for interviewers to be able to choose in any way whether or not they were going to administer a long or short questionnaire for this individual to this household because that certainly introduces biases and a number of data quality issues." KII 2

The relationship between the survey process and data quality is on everyone's mind. Key informants often referenced a conventional wisdom that there is an inverse relationship between the questionnaire length and data quality. Other survey elements, whether the pressure to finish fieldwork quickly, adding content to a questionnaire, or involving more people in the fieldwork, were all referenced as influencing data quality negatively. Several key informants noted that these assumptions were not based on specific evidence, but these relationships were generally accepted truths among those experienced in implementing surveys.

"I mean, nobody has any good evidence about this. So it's just a lot of experienced people who've been doing field work for 30 years on these surveys. And just people knowing how difficult it is to maintain the quality of the data when there's a real push to get the whole survey done." KII 1

"You want to get the most relevant information you can and, particularly with countries wanting to add more and more modules and more and more country-specific questions, if everybody has to answer all of those – you're in trouble in terms of data quality." KII 1

"People say, well, this is the DHS, this is such a high-quality survey. And then you say, well, it won't be high quality if you include all this stuff, but you never win that argument." KII 2

"That if you are increasing these questions like this, content like this, coverage also you are increasing, okay, then quality will be deteriorated completely. How can we trust this quality?" KII 4

"I always think, of the quality. The quality, I mean, if you use more people you need more supervisors. Once the job involves a large number for people it's difficult to control. It's difficult to guarantee everybody will work in the same way…I didn't see a report on the data quality on the last [survey] if it's good or not, I have no idea, but, I will say logically, it's always true. If you involve more people, it's more difficult to finish working, good quality." KII 6

"Or if the questionnaires are too heavy, you try to reduce the content. This is also [a] data quality concern. You reduce the workload of the interviewers; you expect that to improve or to get a better data quality." KII 6

"Normally it's as the sections are at the end of the questionnaire that may be affected more than the section at the beginning of the questionnaire." KII 6

"You could imagine it having a very long questionnaire and dealing about 25 households. I think the interviewers would take too long and also the respondents I think would be fatigued. So, the element of quality would be compromised in terms of maybe the respondents, from the respondents or the interviewer fatigue." KII 7

Some questions included in the questionnaires elicit initial assumptions or feelings about possible data quality issues. Questions that are sensitive or nonstandard are often viewed as inherently problematic and having inferior quality. Including biomarkers is mentioned as increasing quality (response rates) because respondents are often interested in the results of these measures. These issues surpassed the issue whether these questions were located in the long or short questionnaire.

"On the basis of quality that if we are going to collect these data, that will not [be] good for quality… for example, [topic that was eventually included] questions, we didn't want that, that has been imposed on us." KII 4

"The data quality issue is coming, not because of the longer and shorter version of questionnaire, it is coming by nature of the question." KII 4

"People had general perceptions with the increasing number of questions and increasing duration of interview – the data quality will be deteriorating, but our assumption is that since we are taking all biomarkers, [it is] okay." KII 4

"So you find that a situation whereby some interviewer actually will not like to ask the question rather than ask them, but if they're not given an explicit answer, they would rather record their own answer to make them proceed with the interview. So yes, there were some experiences basically in the long questionnaire due to the nature of the questions." KII 8

"… but for sensitive questions, you much more often get an outlier. One interviewer who's an outlier and who's obviously not doing their job properly." KII 1

Key informants are familiar with data quality indicators and can identify quality issues during fieldwork. Key informants do not only discuss data quality in theory, but often discuss specific indicators that they rely on to diagnose data quality issues during fieldwork. Though some informants mentioned long and short questionnaire quality checks, most checks that occurred during fieldwork were not related specifically to long and short questionnaires.

"The suggestion was given [to] me was this: … Don't more overburden them. If you overburden them also, well, … field check tables will be affected. Obviously, if you're overburdening them, there is a huge rejection, or they code every time [the] same thing. And they go further and… it will be showing the percentage of them coming down, a response rate or other indicators also." KII 10

"By reaching Section 10, Section 11 after interviewing for over an hour, there is interviewer fatigue. Some may intentionally [choose] that [the respondent] has never heard of HIV/AIDS. Then they are skipping forty questions." KII 4

"Yeah, look at [this FCT indicator]. Eligible women not interviewed due to lack of privacy. This is a clear example of interviewer fatigue." KII 4

"There is age displacement at upper and lower end [of the eligible age range]. Number of eligible women, men and children, various skip from women's interview. These skips we are sharing… we are knowing those and going back to confirm in the household where we are suspecting. Okay. Or that is why this monitoring this skip is helping us a lot." KII 4

"And if the interviewers are really asking questions in the manner they are supposed to ask. If they are asking all the questions in all the modules, in the long and the short, it was really good." KII 7

"I don't know that we included any on [long and short questionnaires] specifically, I don't think so." KII 5

"… that addresses the data quality, to see if [long and short questionnaires] are randomly assigned… no one has the opportunity to look at a household and say, I'm going to do a long here, but then it also has a number of data quality checks, we can see that the fields are appropriately doing the long and short questionnaire, because we know which households that have completed that questionnaire." KII 2

4.1.6 Opinion and future

At the beginning stages of the survey process, key informants noted that there were reservations about having long and short questionnaires. One key informant also discussed the long questionnaire and their unease about what issues may arise from having such a long questionnaire.

"The steering committee individuals may not have seen the need for [long and short questionnaires] in the beginning; by the end of the survey, they thought it was a useful idea." KII 2

"When I started the training, I saw the questionnaire and it was so huge and I was shocked. So people are ready to answer these, all these questions?" KII 10

"So really is it feasible to talk with a woman for more than two hours? … There will be practical problems. There will be someone coming inside and saying how much time?... Some people who are experienced, they were saying that, yeah, these are the problems [that] will come up. We have [to] tell them … 'Okay, up to here I completed, sister I'll be coming after sometime so that we will be covering the remaining parts'. So there was a discussion about this long questionnaire." KII 10

The long and short questionnaire format is generally regarded as valuable. The approach of using long and short questionnaires was well received by stakeholders and the IAs involved.

"I think it's a great idea. I think everybody should do it if they've got a large enough sample size for the reporting requirements. … I mean, why collect a lot of data that you don't need really?" KII 1

"I don't think anybody expressed any negative thoughts about having the long and short questionnaire." KII 1

"I think I sort of accepted it as, this is the best idea and this as a really solid, reasonable, solution. That opinion did not change over time." KII 2

"Everyone appreciated that. Because the issues we are covering in [the] long questionnaire, those issues are not required to be estimated at a [lower administrative unit] level. People are satisfied with [upper administrative unit] level estimates of those issues, that is why everyone appreciated this." KII 4

"This is an interesting solution to the challenge of everyone wanting all the data." KII 5

"I would say it's useful also because it also helps us now to cover, capture some indicators that are at closer levels without the fatigue or compromising the quality because, you know, if you have the mixture of the two, then you try to cut on the fatigue from the enumerators." KII 8

"It may not be relevant to have [all indicators] at a [lower administrative] level, and if we needed them, then it would mean we had to have an expanded sample size. So it would definitely have saved on time, resources, and everything. So, this splitting up the [administrative units] or splitting up the questionnaires in the [long and short questionnaires] – I think that was a workable and a good idea. Yeah, so I think it definitely is a good strategy." KII 9

When asked if long and short questionnaires were a useful feature for future surveys, every key informant agreed this approach was beneficial. The informants discussed using long and short questionnaires to solve issues with demands for longer surveys, nonstandard age populations, and lower-level estimates. Long and short questionnaires also offered positive effects like reducing cost and overall survey burden.

"Yes. Where you're doing nonstandard age populations, there's no reason to repeat the programming for the questions that you definitely want to link to the additional data. So from that standpoint, yes." KII 3

"I would say so. The idea of having short and long questionnaires or having modules for subsamples in general sounds like an efficient solution. Collect the necessary data only. It has implications on the cost and on the burden in general… I mean, the challenge before with paper [questionnaires] was how to train interviewers on dealing with these, complex designs. I believe, from my humble experience in the field, with CAPI, I believe that should be easier." KII 5

"Yes... It is a useful tool to get more estimates on indicators at the lower geographical level." KII 7

"Yes. I feel that it is very useful… it reduced the cost…This is a very smart thing to do…. I feel [long questionnaire] had a good impact in our survey." KII 10

Deploying long and short questionnaires has implications for future surveys: it may set a precedent or make it easier to find solutions for future surveys. After the first experience with a survey that uses long and short questionnaires, both country stakeholders and key informants consider them as a useful option (if not a default approach) for survey implementation.

"I would say [the decision to have long and short questionnaires] was largely based on the previous survey. What the Ministry had decided was that the [previous survey] survey would be considered the baseline for future surveys. And therefore they tried to stick pretty closely with what was in that survey. So to facilitate looking at trends over time." KII 1

"… for any survey in this country, [first survey to use long and short questionnaires] should be treated as benchmark. It means if you will have to go for same design, same protocol. So that comparability can be maintained." KII 4

"As a matter of fact, when we have a nonstandard design, and once I manage to fulfill or complete any of the implications of these nonstandard designs on my job, you know, I'll be happy because… when you do such designs and if it has implications on how you do your work, this doesn't mean that you shouldn't do it next time. It is actually, you have to adapt, you have to change your tools, which actually on the long run, it works, it is worth it because you will find that another [survey] had a similar situation. So now it's easy for you to use a tool that you previously customize and customize for the new [survey] and to do the job." KII 5

"And we are also going into the [upcoming survey]. I don't know what format will be used, but I hope maybe assuming that methodology will be used because now we still want the [lower administrative unit] indicators." KII 7

5 DISCUSSION

5.1 Data Quality Indicators

This report used mixed methods to investigate data quality and implementation differences in long and short questionnaires. The quantitative section utilizes indicators rarely discussed in the literature but often used in the monitoring of fieldwork, as well as indicators that have often been used in post hoc analyses to assess data quality to search for intentional workload reduction by interviewers. The long questionnaires in each country had large differences in the average number of variables per woman compared to a country's short questionnaires. In Kenya, the long questionnaire lasted twice as long as the short questionnaire on average. Despite these differences, there is little evidence that interviewers made different efforts to reduce their workloads based on whether they were administering the long or the short questionnaire. There is evidence of upward age displacement from ages 49 to 50 in India and Kenya and downward age displacement from age 15 to 14 in Kenya, yet no evidence that age displacement differs between long and short questionnaires within the same country. There was no evidence of interviewers intentionally skipping sections, proxied by significant differences in filter questions between long and short questionnaires. Completeness of date of birth for both women age 15-49 and live births in the last 5 years were extremely similar between long and short questionnaires in their respective countries, even when that difference was statistically significant as exhibited in Kenya.

5.2 Key Informant Interviews

In a series of key informant interviews with survey experts who implemented DHS surveys with a long and short questionnaire design, we identified five themes that describe the key informants' experiences and opinions of long and short questionnaires.

Deploying long and short questionnaires in order to obtain estimates of a subset of survey indicators for nonstandard populations or at lower administrative levels solves a significant problem in survey design and implementation. It does not vanquish all challenges, but it is widely embraced as a useful approach to survey design.

The major problem it solves is meeting in-country data demands while maintaining feasibility. By definition, survey stakeholders are people who are interested in data, and the point of a survey is after all to obtain data for users. Stakeholders see DHS surveys as an opportunity to obtain high-quality data and advocate vehemently for the inclusion of their topics or populations of interest. The pedestrian realities of funding and implementation complexity can make these appetites difficult to satisfy within funding constraints or without creating implementation challenges. It is rare for any one funder to be able to cover the entire cost of a survey, and an absolutely herculean degree of sustained effort, concentration, and attention to detail on the part of many individuals – not least fieldworkers – is required to successfully implement even the most modest DHS survey. A larger sample size or a significantly longer questionnaire requires either more fieldworkers or a longer duration of fieldwork. These factors both increase cost and require a more complicated management structure (e.g., to assure high-quality training, to manage fieldwork logistics, to conduct data quality monitoring during fieldwork). The longer fieldwork is, the bigger the risks to the health and safety of fieldworkers (for example if they have to work through certain

seasons, or spend more time in areas with security concerns). Using long and short questionnaires can in fact minimize the unwieldiness of implementation. While a survey with long and short questionnaires may mean a more complex design, it is, somewhat paradoxically, less onerous to manage; and the now common use of CAPI makes the deployment of the different questionnaires seamless for fieldworkers. A survey that is easier to manage should be a survey in which more attention can be paid to data quality. This maximizes what is possible within the available budget.

Of course, challenges remain. First, 'easier to implement' is not the same thing as 'easy to implement', and there are still concerns about data quality. The more complex design is not easier from a data processing perspective – it increases the time needed for programming, and every divergence from standard DHS CAPI programming increases the possibility of error. It remains a general assumption that longer questionnaires are more fatiguing (for both interviewers and respondents), and that even if they are implemented with fewer respondents and even if quality is consistent between both questionnaires for the common content, that quality may drop off for the content that is only in the long questionnaire. This may particularly be an issue for any country-specific or nonstandard questions that have not been validated, on which it can be more difficult to provide high-quality training. Second, implementing long and short questionnaires is no guarantee that everyone will be happy. For survey stakeholders, some questions will still not be included, some results may be disappointing, and their understanding of what data will be available at the end may be imperfect. Secondary users may have a difficult time understanding the survey design, or why the data may not be suitable for all the analyses they desire to carry out.

However, key informants were unanimously supportive of the concept. Key informants played various roles on surveys and brought very different perspectives to bear, but they all thought of using long and short questionnaires as an effective way to meet data users' needs while keeping survey implementation feasible and survey data quality high.

Another possible benefit of the way surveys with long and short questionnaires are often designed is that the inclusion of biomarker collection in the 'long' households may make the lengthier household and/or individual interviews more palatable to respondents. The design of surveys that include long and short questionnaires often looks lopsided, with many or all of the 'extras' (long household questionnaire, long woman's questionnaire, man's questionnaire, and all or supplementary biomarkers) taking place in the same households. While survey implementers' initial instinct may be to distribute aspects of the survey more equally across households, this would reduce the usefulness of data; the domestic violence module is a case in point. The domestic violence module shifted from standard implementation in households where men are not interviewed to those households where men are interviewed, to facilitate analysis of domestic violence respondents and their husbands together. This unequal distribution may not in fact be particularly noticeable to respondents to the household and woman's questionnaire (who may not discuss details of content with neighbors in differently selected households), but biomarkers are a source of interest to respondents and their communities. The results of anthropometry and tests with point-of-care results are

shared with respondents, and some outcomes result in referral to health care or immediate treatment.[7] This is generally perceived positively by respondents and their communities; it is common for households not selected for the survey or not selected for biomarkers to be upset about their exclusion on this basis. The fact that households selected for the 'long' questionnaire are more likely to either have biomarker collection at all or have supplementary biomarker collection may reduce respondent impatience with the length of the questionnaire – although again, the respondents are unlikely to know that there are different versions of questionnaire.

5.3 Limitations

This report has several limitations. This analysis did not attempt to make summary conclusions of the data quality of each of these surveys according to a threshold or standard, rather we only discussed if there are differences between the long and short questionnaire in each survey. For this reason, the analysis does not stratify by elements that often reveal data quality issues, such as interviewing teams or fieldwork time, and though our analysis reveals no data quality differences between the long and short questionnaires in the three surveys, this does not suggest that there are no data quality issues overall that could be related to questionnaire length. Moreover, we cannot determine if there are data quality issues in data derived from questions only asked in the long questionnaire because we cannot compare these data exclusive to the long questionnaire to data from the short questionnaire. Further, some of these data quality indicators may be used in fieldwork but are only proxy measures that try to estimate phenomena that are difficult to capture. For example, when examining whether sections of the questionnaire are skipped during an interview, it is important to note that these estimates may be accurate data and not a reflection of intentional skipping. Theoretically, these estimates should be the same between long and short questionnaires that are sampled equally in the same clusters. However, because we do not examine lower-level estimates of these indicators (e.g., by team or interviewer), it is possible that differences exist between long and short questionnaires that are masked at a national level.

Our key informant interviews were limited by a small sample size (n=10) and each survey was not represented by an equal number of informants. Informants did not include data processing staff; this skewed our qualitative analysis towards the CAPI user experience over the substantial programming burden on data processing staff in surveys with long and short questionnaires. Moreover, these surveys took place a long time ago, and recall may be poor and affect our key informants disproportionately. Additionally, because our interviews centered around data quality, informants may have been affected by a social desirability bias and avoided describing negative data quality issues. However, it was generally felt that our informants were candid and open about their experiences. Finally, our survey selection for this report was not methodical and we were hindered by the lack of authoritative definition of what constitutes long and short surveys. There is no easy way to search through DHS datasets to estimate the number of questions or modules that

[7] When respondents (usually children age 6 months to 5 years and women age 15-49 years) have low hemoglobin results, a referral slip is provided to take to a health facility to seek treatment. The standard population for malaria testing is children age 6 months to 5 years and in the case of malaria rapid diagnostic tests, the country first-line treatment is offered when the result is positive, providing the child is not already receiving treatment and does not either have a low hemoglobin result or recent symptoms of severe malaria. When a child has malaria and low hemoglobin, or malaria and recent symptoms of severe malaria, that child is referred for treatment.

were included, and it is possible that there were other surveys with long and short questionnaires that would have added value to this report.

5.4 Recommendations

The use of long and short questionnaires should be included in the survey design options for DHS surveys (see all recommendations in Table 11). In order to evaluate the inclusion of long and short questionnaires in a survey, survey designers need to be equipped with tools to more accurately estimate the impact of this design on the duration of fieldwork and on the survey budget. In terms of questions included in the long and short questionnaires, it remains unclear whether questions that are difficult to train on and implement should be placed in the long or short version. One KI recommended placing more difficult questions, for example questions where the respondent has to consider a hypothetical situation, in the long questionnaire as it would be implemented less often. On the other hand, another KI mentioned that interviewers were much more comfortable asking questions in the short questionnaire as they had repeated it so often. Based on which questions are included in which questionnaire, tabulations for the survey final report should be developed at the beginning of the survey process to ensure any change in denominators for indicators of interest are clearly explained early on and survey stakeholders are not surprised at the report writing or final report stage.

Whether or not interviewers should be blinded as to whether their interviews with respondents will be using the long or short questionnaires is a topic for further consideration. There are three recommendations for fieldwork implementation. The first is to have teams begin their work in a given cluster with the long questionnaire interviews so they can have adequate time to review and edit them before their departure from the cluster. The second is to equitably distribute the long and short questionnaires among interviewers to prevent overburdening some and causing feelings of unfairness. The third is, if some interviewers are weaker at the beginning of fieldwork and feel more comfortable with the short questionnaire, to develop a process for coaching them to be able to administer the long questionnaire with skill and confidence. In one survey, stronger interviewers used this coaching technique to build up the capacity of their weaker teammates. Key informants generally felt that the standard FCTs were adequate for monitoring fieldwork for a long and short questionnaire survey. Further, our quantitative results show that even if there were special FCTs developed for this survey design, there would be little to no difference by questionnaire reflected in the tables.

Table 11 Recommendations for future surveys

Survey phase	Recommendation
Survey and sample design	Include long and short questionnaire approach in the survey design options.
Survey and sample design	Tabulations should be developed at the beginning of the survey process for long and short questionnaires (or other country-specific subsampling) and shared with stakeholders
Fieldwork implementation and monitoring	Structure fieldwork around length of questionnaires
Fieldwork implementation and monitoring	Long and short questionnaires should be equitably distributed among interviewers and interviewers should be coached to implement long questionnaires with confidence

5.5 Conclusion

In summary, this report analyzed data from three DHS surveys that utilized long and short questionnaires – the 2016 South Africa DHS, 2014 Kenya DHS, and 2015-16 India NFHS – and sought experts who implemented those surveys to provide insight into their experiences with long and short questionnaires. We found little evidence that having differing lengths of questionnaires resulted in data quality differences between the resulting data from the two questionnaires. Key informants agree that deploying long and short questionnaires in order to obtain estimates of a subset of survey indicators for nonstandard populations or at lower administrative levels solves a significant problem in survey design and implementation. Key informants were unanimously supportive of using long and short questionnaires in the future. We recommend that the use of long and short questionnaires be included in the survey design options for future DHS surveys so that surveys can meet in-country data demands while maintaining feasibility.

REFERENCES

Allen, C. K., T. N. Croft, T. W. Pullum, and S. M. L. Namaste. 2019. *Evaluation of Indicators to Monitor Quality of Anthropometry Data During Fieldwork.* DHS Working Paper No. 162. Rockville, Maryland, USA: ICF. https://dhsprogram.com/pubs/pdf/WP162/WP162.pdf.

Assaf, S., M. T. Kothari, and T. W. Pullum. 2015. *An Assessment of the Quality of DHS Anthropometric Data, 2005-2014*. DHS Methodological Reports No. 16. Rockville, Maryland, USA: ICF International.

Bradley, S. E. 2016. *When Quality Matters: Linking the Reliability of Demographic and Health Survey Data to Biases in International Mortality, Fertility, and Family Planning Estimates*. Unpublished Doctoral dissertation, UC Berkeley.

Hill, K., K. Thomas, C. AbouZahr, N. Walker, L. Say, M. Inoue, and E. Suzuki. 2007. "Estimates of Maternal Mortality Worldwide between 1990 and 2005: An Assessment of Available Data." *The Lancet* 370 (9595): 1311-1319.

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.

Lyons-Amos, M., and T. Stones. 2017. "Trends in Demographic and Health Survey Data Quality: An Analysis of Age Heaping over Time in 34 Countries in Sub Saharan Africa between 1987 and 2015." *BMC Research Notes* 10 (1): 760. https://www.ncbi.nlm.nih.gov/pubmed/29262857.

Perumal, N., S. Namaste, H. Qamar, A. Aimone, D. G. Bassani, and D. E. Roth. 2020. "Anthropometric Data Quality Assessment in Multisurvey Studies of Child Growth." *American Journal of Clinical Nutrition*. https://www.ncbi.nlm.nih.gov/pubmed/32672330.

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

Pullum, T. W. 2019. *Strategies to Assess the Quality of DHS Data*. DHS Methodological Reports No. 26. Rockville, Maryland, USA: ICF.

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.

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. https://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.

Short Fabic, M., Y. Choi, and S. Bird. 2012. "A Systematic Review of Demographic and Health Surveys: Data Availability and Utilization for Research." *Bulletin of the World Health Organization* 90 (8): 604-12. https://www.ncbi.nlm.nih.gov/pubmed/22893744.

Stanton, C., N. Abderrahim, and K. Hill. 2000. "An Assessment of DHS Maternal Mortality Indicators." *Studies in Family Planning* 31 (2): 111-23. https://www.ncbi.nlm.nih.gov/pubmed/10907277.

APPENDIX I

Key Informant Interview Guide

Introduction and consent

I have gotten in touch with you because we are writing a report analyzing the data quality in DHS surveys that used long- and short-form questionnaires. The purpose of the study is to assess if there are differences in data quality between long- and short-form questionnaires by analyzing data quality indicators. Another aim of this study is to understand other implications of administering long or short questionnaires through key informant interviews with survey implementation staff such as yourself who implemented both long- and short-form questionnaires. The report will be published and shared both internally within The DHS Program and externally to a wider audience. Thank you for agreeing to talk with us today to share your experience and recommendations. Our discussion today will be focusing on a few core survey phases, including:

- Survey Design
- Survey Content
- Sample Design
- Fieldworker Training
- Fieldwork and Fieldwork Monitoring

[Interviewer circle or bold ahead of time the phases that will be covered.]

Your participation in this interview is completely voluntary. If you want to stop at any time or don't feel comfortable answering a question, please let me know. Taking part in this interview is your agreement to participate.

This discussion will be recorded, and the recordings will be kept securely and only accessed by the research team. We would also like to collect demographic information about the experts we have interviewed. We will send you a separate email with a code and a link to a brief demographic survey online. Please fill that out when you have the time.

The information you provide as part of the interview and the survey will not be attributed to you individually but your name will appear in the appendix of the report. Do I have your permission to record the discussion and include your name in the appendix of the report? Our discussion today will take [45-90 minutes - select duration based on sections of interview that will be administered]. Do you have any questions before we start? [Answer any questions.]

Let's begin. At the beginning, I am going to ask a few basic questions about your background, but we will quickly move on to your experience with and views about different survey phases. Please feel free to respond to questions that you are not an expert in if you have something to share. [START RECORDER]

Appendix Table 1 Interview Guide for Key Informant Interviews

No.	Question	Skips for each respondent's interview				
		Survey designer	Survey manager	Sampler	Implementing agency staff	Fieldworker
Role		**1**	**1**	**1**	**1**	**1**
1	During this conversation, we will be discussing the (survey, country, year). Can you describe your role on that survey?					Skip to Q.22
Involvement in decision to have short and long questionnaires		**7**	**7**	**7**	**7**	**0**
2	Were you involved in the decision-making about having short and long questionnaires?	If No, Skip to Q.8	If No, Skip to Q.8	If No, Skip to Q.8	If No, Skip to Q.8	
3	Could you describe how that decision was made?					
4	What factors were considered?	If No Mention of Data Quality, Skip to Q.6	If No Mention of Data Quality, Skip to Q.6	If No Mention of Data Quality, Skip to Q.6	If No Mention of Data Quality, Skip to Q.6	
5	Please talk more about how data quality was part of the decision.					
6	Could you describe how, if at all, data quality was part of the decision?	Check Q.4. If Mentioned Fieldwork Duration, Skip to Q.9	Check Q.4. If Mentioned Fieldwork Duration, Skip to Q.9	Check Q.4. If Mentioned Fieldwork Duration, Skip to Q.9	Check Q.4. If Mentioned Fieldwork Duration, Skip to Q.9	
7	Was the duration of fieldwork part of the decision? If so, could you talk about how?	Skip to Q.9	Skip to Q.9	Skip to Q.9	Skip to Q.9	
8	What is your understanding of why the decision was taken to have short and long questionnaires?					
Content of long and short questionnaires		**7**	**7**	**7**	**7**	**0**
9	Were you involved in the decision-making about what content to have in the short and long questionnaires?	If No, Skip to Q.15	If No, Skip to Q.15	If No, Skip to Q.15	If No, Skip to Q.15	
10	What other parties were involved in the decision-making about what content to have in the short and long questionnaires?					
11	Could you describe how that decision was made?					
12	What factors were considered?	If No Mention of Data Quality, Skip to Q.14	If No Mention of Data Quality, Skip to Q.14	If No Mention of Data Quality, Skip to Q.14	If No Mention of Data Quality, Skip to Q.14	
13	Please talk more about how data quality was part of the decision.	Skip to Q.16	Skip to Q.16	Skip to Q.16	Skip to Q.16	
14	Could you describe how, if at all, data quality was part of the decision?	Skip to Q.16	Skip to Q.16	Skip to Q.16	Skip to Q.16	
15	What is your understanding of how the content of the short and long questionnaires was decided?					

Continued…

Appendix Table 1—*Continued*

No.	Question	Skips for each respondent's interview				
		Survey designer	Survey manager	Sampler	Implementing agency staff	Fieldworker
Sampling of short and long questionnaires		**6**	**6**	**6**	**6**	**0**
16	Were you involved in the sample design and decisions about how to distribute the short and long questionnaires?	If No, Skip to Q.21	If No, Skip to Q.21	If No, Skip to Q.21	If No, Skip to Q.21	
17	Could you describe how the sampling decisions were made?					
18	What factors were considered?	If No Mention of Data Quality, Skip to Q.20	If No Mention of Data Quality, Skip to Q.20	If No Mention of Data Quality, Skip to Q.20	If No Mention of Data Quality, Skip to Q.20	
19	Please talk more about how data quality was part of the decision.	Skip to Q.22	Skip to Q.22	Skip to Q.22	Skip to Q.22	
20	Could you describe how, if at all, data quality was part of the decision?	Skip to Q.22	Skip to Q.22	Skip to Q.22	Skip to Q.22	
21	What is your understanding of how the sample was designed, particularly how to distribute the short and long questionnaires?	Skip to Q.37		Skip to Q.37		
Training on short and long questionnaires		**0**	**4**	**0**	**4**	**4**
22	Were you involved in planning or executing interviewer training?					
23	How were the short and long questionnaires covered during the training?					
24	Did the fact of there being short and long questionnaires make anything about the training different? If so, how?					
25	Was there explicit discussion during the training (for interviewers, supervisors, or both) about the possible effects on data quality of having short and long questionnaires? If so, please talk more about what was discussed.		Skip to Q.32		Skip to Q.32	
Short and long questionnaire fieldwork		**0**	**0**	**0**	**0**	**6**
26	Could you talk about how having both short and long questionnaires affected actually doing fieldwork?					
27	Did it affect how you distributed assignments to your team? If so, how?					
28	Did team members have opinions about the short and long questionnaires? If so, what were their opinions?					
29	Did they have different experiences using the different questionnaires? If so, what were their experiences?					
30	Do you think the different lengths of the questionnaires affected how they asked questions, probed, or recorded responses? If so, please talk more about how the questionnaire length affected the interviews.					
31	Did respondents notice or ask about there being two different questionnaires? If so, what do you remember the respondents noticing or asking about?					Skip to Q.37

Continued…

Appendix Table 1—*Continued*

No.	Question	Skips for each respondent's interview				
		Survey designer	Survey manager	Sampler	Implementing agency staff	Fieldworker
Fieldwork monitoring		**0**	**5**	**0**	**5**	**0**
32	Were any of the field check tables affected by having short and long questionnaires? If so, how were they affected?					
33	Did you consider the short and long questionnaires in your interpretation/use of the field check tables? If so, how were they considered?					
34	Did you observe any fieldwork (practice or main)?		If No, Skip to Q.37		If No, Skip to Q.37	
35	What did you notice about the implementation of the short and long questionnaires?					
36	Did you notice anything that raised concerns about data quality? If so, please talk more about the concerns.					
Opinion of short and long questionnaires		**4**	**4**	**4**	**4**	**4**
37	At the beginning of your involvement with the survey, what assumptions did you have about how having a short and long questionnaire would play out?					
38	Did your opinion change as the survey progressed?					
39	Do you think this is a useful feature to consider using in other surveys?					
40	Why or why not?					
India-specific question		**1**	**1**	**1**	**1**	**1**
42	Both the 2015-16 and 2019-20 NFHS surveys had short and long questionnaires. Did you make any changes in design, sampling, questionnaire content, training, or any other aspect of the survey as regards the short and long questionnaires for the 2019-20 round based on lessons learned during the 2015-16 round?					
Closing questions		**2**	**2**	**2**	**2**	**2**
43	Is there anything we haven't covered that you'd like to share about the short and long questionnaires? We are particularly interested in issues of data quality.					
44	Do you have any questions for me?					
COUNT OF MAXIMUM NUMBER OF Qs PER RESPONDENT		**28**	**37**	**28**	**37**	**18**

APPENDIX II

List of Key Informants

Dr. Fred Arnold
Mahmoud Elkasabi
Mani Deep Govindu
Sara K. Head, PhD, MPH
Andrew Amina Imbwaga, (HSC), MSc (Demography)
Joanna Lowell
Zachary O. Ochola
Dr. Ruilin Ren
Professor S.K. Singh
Yelamanchili Vaidehi