Endogenous Sample Selection in Household Surveys*

Torsten Figueiredo Walter† Niclas Moneke‡

-PRELIMINARY AND INCOMPLETE-
-PLEASE DO NOT CIRCULATE-

January 15, 2024

This paper examines the production of household survey data, a crucial input to social science research and evidence-based policy. In theory, household surveys generate data from a randomly selected, representative sample of the population. We provide causal evidence that in practice, enumerators respond to the design of surveys by excluding household members that require a disproportionate amount of effort on their end. Leveraging 146 Demographic and Health Surveys and 52 Multiple Indicator Cluster Surveys from 79 low- and middle-income countries, we show that household members eligible for large numbers of questions are frequently underrepresented in survey samples. Omitted members differ systematically from included members and as a result, survey samples are not representative of the population, thus leading to biased aggregate statistics.

*We thank Claudia Alcarra, Mouhamed Ba, Ana Barros, Frederic Cochinard, Ameni Hajji, Abigail Lamptey, Sophia Lin, Shifa Muhammad, Alfred Shauri, Jordan Simpson, Kevishen Valeyatheepillay, Emma Wang and Ruike Zhang for research assistance. Financial support from STEG is gratefully acknowledged.

†New York University Abu Dhabi, email: t.f.walter@nyu.edu

‡University of Oxford, email: niclas.moneke@economics.ox.ac.uk

1 Introduction

Household survey data is a crucial input to empirical research in the social sciences as well as evidence-based policy design. Therefore, it is important to understand the underlying data generating process and its implications for data quality. In theory, household surveys generate data from randomly selected, representative samples of the population. In practice, enumerator incentives may not be aligned with representative random sampling, however. A simple model of enumerator behavior highlights the incentive to avoid high-effort survey subjects if these can be identified ex-ante and the expected cost of avoidance is low. Thus, it is important to assess empirically to what extent enumerator incentives lead to endogenous sample selection and what this implies for statistical inference.

This paper starts from the observation that all household surveys generate variation in the expected enumerator effort costs across household members by limiting the eligibility for specific (sets of) questions to specific household members. The 2006 MICS in Togo provides an illustrative case. Figure 1 highlights how household members that are eligible for long individual questionnaires (woman's questionnaire for women aged 15 to 49 and under five questionnaire for children below 5) are selected based on the roster, thereby making effort cost differences between household members of different age and gender very salient. Figure 2 displays the associated distribution of the average number of questions asked about household members by age and gender, with a particularly high question volume for children below the age of 5 and women between the ages of 15 and 49 (top panels). This creates an incentive for enumerators to avoid such high-effort household members, by omitting them from the roster or manipulating their demographics, in particular gender and age, such that they are not eligible for individual questionnaires. In fact, the bottom panels of Figure 2 show how the associated age distributions lack mass in all age ranges that are eligible for individual questionnaires and have excess mass on the ineligible side of eligibility thresholds.

In this paper, we document more broadly how variation in question load across household members leads to endogenous selection of household members into survey samples and study its implications for statistical inference. Leveraging 146 Demographic and Health Surveys and 52 Multiple Indicator Cluster Surveys from 79 low- and middle-income countries, we provide causal evidence that enumerators frequently screen out household members eligible for individual ques-

tionnaires. To this end, we employ two complementary empirical strategies. First, we exploit the random assignment of a man's questionnaire to a subset of households in 184 DHS and MICS to identify the effect of additional questions on the recording of men of eligible age in household rosters. Second, we adopt a difference-in-differences strategy comparing the recording of women of eligible age to the recording of men of eligible age in the DHS/MICS relative to a contemporaneous population census, restricting our attention to DHS and MICS without a man's questionnaire. This means that in the considered surveys women of eligible age face many more questions than men of the same age while there is no difference in question load between these two groups in the population census.

We find that the eligibility households for individual questionnaires leads to a significant reduction of members of eligible age in two thirds of surveys, with the reduction in household members ranging between 2% and 21% across surveys. The extent to which this is counteracted by an increase in the number of ineligible household members through age displacement varies greatly. In some surveys, age displacement can explain all of the missing household members. In other surveys, it cannot explain any of it, suggesting that the missing household members have simply been omitted from rosters.

Missing household members differ systematically from included ones. By comparing men of eligible age in households that are eligible and ineligible for the man's questionnaire, we show that they are often younger, less educated and less closely related to the head of their household.

This selection renders samples unrepresentative and leads to bias. Individual-level statistics, such as the national average educational attainment, are distorted in many surveys. Similarly, household-level statistics normalized by household size are biased because the endogenous omission of household members leads to a reduction in observed household size.

This paper contributes to three streams of literature. First, it adds to an old literature on selection in surveys (see Mogstad et al. 2022 for a recent contribution). While this literature is largely focused on non-response bias, i.e., self-selection of respondents, this paper highlights another margin of selection, namely the screening of respondents by enumerators. Second, this paper contributes to a nascent literature on enumerator behavior (LoPalo 2023) by demonstrating how survey design and implementation protocols shape enumerator incentives and thereby affect the recording of household members. Third, it is related to a recent stream of work on household

survey design and respondent fatigue (Ambler et al. 2021, Aggarwal et al. 2023). In contrast to this body of work, though, it focuses on the effect of question load on enumerator behavior rather than respondent behavior.

The remainder of the paper is structured as follows. Section 2 provides background information on the data used in this paper and how it is generated. Section 3 outlines the two empirical strategies used to identify the effect of question load on the screening out of household members. Section 4 presents the main results and section 5 provides additional evidence on the selection of household members out of sample. Section 6 discusses implications for aggregate statistics. Finally, section 7 concludes.

2 Background

2.1 Demographic and Health Survey and Multiple Indicator Cluster Survey

In this paper, we study endogenous sample selection in two large international household survey programs, the Demographic and Health Survey (DHS) and the Multiple Indicator Cluster Survey (MICS). The DHS focuses on fertility, family planning, maternal and child health, gender, HIV/AIDS, malaria, and nutrition. It is funded by USAID and implemented by ICF. The MICS focuses on the situation of children and women and is supported by UNICEF. The former program started in 1984 while the latter began in 1990s. Both programs have a reputation for collecting accurate, comparable, nationally representative data using standardized survey instruments across countries.

We focus on these household survey programs for three reasons. First, they are of great relevance for research and policy. The DHS and MICS are commonly used data sources in empirical social science research. In the policy realm, each of the two programs is the basis for about 30 SDG indicators. Moreover, aid flows have been explicitly conditioned on indicators derived from them (e.g., World Bank Program-for-Results). Second, the global coverage of low- and middle-income countries by both programs alleviates concerns related to external validity. Since program inception, more than 400 DHS have been conducted across 90 countries. Over the same time period the MICS has covered 120 countries in more than 350 surveys. Third, the random assignment of the man's questionnaire to households provides us with a source of exogenous variation in the effort cost

associated with men of eligible age.

In this paper, we focus two subsets of surveys from these programs. First, we study all 184 surveys in which a random subset of household was assigned to have a man's questionnaire administered. Second, we examine 14 surveys in which at least a subset of households did not receive a man's questionnaire and for which microdata from a contemporaneous population census is available. Overall, our analysis includes 198 surveys across 79 countries. Figure 3 illustrates the geographic coverage of the data.

2.2 Survey design and implementation

USAID/ICF and UNICEF provide questionnaire templates to local agencies at the beginning of each survey wave. The DHS originally consisted of two questionnaires: a household questionnaire (including household roster) and a woman's questionnaire. The MICS was originally composed of three questionnaires: a household questionnaire (including household roster), a woman's questionnaire and an under-five questionnaires. In both survey programs, the household questionnaire is composed of two parts, the household roster and household-level questions. The household roster gathers basic demographic information on all household members and is used to determine the eligibility of household members for individual questionnaires based on gender and age. Household-level questions concern topics such as asset ownership, energy use and sanitation. The woman's questionnaire is administered to all women aged 15 to 49 and focuses on fertility and maternal health. The under-five questionnaire is administered to all children under the age of 5 and focuses on child health and development.

In later survey waves, both survey programs introduced a man's questionnaire, asking about fertility, health and sexual behavior, with the eligible age typically ranging from 15 to 49 (in some cases also 15 to 54, 59 or 64). Importantly, this questionnaire is not administered in all sample households in all surveys. Instead, in many surveys it is only administered to a random subset (most commonly one half or one third) of households within each enumeration area.

Individual questionnaires are administered after the household roster has been completed. This implies that at the time of the roster completion, survey respondents do not know how the age and gender of household members recorded in the roster affect the length of the household interview. Enumerators are very much aware of this, however, since they are familiar with the survey

structure from their training and their experience with previous households. Moreover, the survey instruments make the eligibility of household members for individual questionnaires very salient, asking enumerators to mark every eligible member as they fill in the roster (see Figures 1 and 5 for illustrations).

2.3 Enumerator incentives

DHS and MICS are funded and supported by USAID and UNICEF, respectively. However, they are implemented by local agencies, most commonly National Statistical Offices. While enumerator contracts vary across agencies, temporary contracts with fixed pay per day are standard. The daily workload of enumerators is typically set in advance by the central office of the implementing agency and adherence to the schedule is heavily emphasized during fieldwork. This means that shorter interviews imply shorter working days for enumerators.[1]

2.4 Data harmonization

While DHS data is accessible in a harmonized format from the website of the DHS program, MICS data is not harmonized across surveys. We thus harmonize all 52 MICS datasets used in this paper to allow for comparison with the DHS.

3 Empirical strategy

In this paper, we employ two different empirical strategies. First, we exploit the random assignment of the man's questionnaire to a subset of households in DHS/MICS to estimate the causal effect of the question load of men of eligible age on the recording of men in household rosters. Second, we adopt a difference-in-difference strategy comparing the recording of women and men of eligible age in contemporaneous DHS/MICS and population censuses to estimate the effect of the woman's questionnaire on the recording of women in survey rosters.

[1] See LoPalo (2023) for more details on the implementation of the DHS.

3.1 Random assignment of man's questionnaire

We identify 184 DHS and MICS carried out across 72 countries between 1991 and 2022 in which a man's questionnaire was administered to a random subset of households. In all surveys, randomization was stratified by enumeration area, with the treatment probability varying between 1/7 and 2/3 across surveys.

The eligibility of households for the man's questionnaire is salient to enumerators. Before recording the household roster, enumerators mark whether the household at hand is eligible for the man's questionnaire (see Figure 4 for illustration). Moreover, in the roster they only mark men as eligible for the individual questionnaire if the household at hand is eligible for the man's questionnaire (see Figure 5 for illustration).

The man's questionnaire typically covers the following topics: marriage, sexual behavior, fertility, domestic violence and HIV. Sometimes it also includes sections on the use of media, ICT, alcohol and tobacco. The median duration of the man's questionnaire varies between 6 and 62 minutes across surveys.

Relying on the random assignment of the man's questionnaire, we run the following OLS regression for each survey in our sample:

$$Y_{ic} = \alpha_c + \beta MQ_{ic} + \epsilon_{ic} \tag{1}$$

where Y_{ic} is an outcome of interest of household i in stratum c, MQ_{ic} is an indicator for the man's questionnaire being administered and α_c is a set of stratum fixed effects. In most surveys, strata correspond to enumeration areas. In a few MICS, the random assignment of the man's questionnaire is additionally stratified by the presence of children below the age of 5 during the enumeration area listing exercise preceding the survey.

3.2 Difference-in-difference with population census

To estimate the extent to which women eligible for the woman's questionnaire are missing from DHS/MICS samples, we adopt a difference-in-difference strategy comparing the recording of women and men of eligible age in DHS/MICS household rosters to their recording in contemporaneous

population censuses. For this purpose, we restrict our sample to two types of DHS/MICS: (i) surveys that do not include a man's questionnaire at all, and (ii) surveys that only administer a man's questionnaire in a subset households. In all of these surveys women of eligible age are subject to many more questions than men of eligible age (in ineligible households). In population censuses, on the contrary, question loads are very similar for women and men of eligible age.

We form 14 pairs between DHS/MICS and contemporaneous population censuses for which microdata is available from IPUMS-International or national data archives (see Table 1).[2] We scale the relative sample weights from the DHS/MICS such that the total number of households in each survey and the paired population census is equal. Figure 7 demonstrates the difference in female-to-male question ratio between DHS/MICS and population censuses. In the two survey programs, the question load of women of eligible age is 3 to 12 times higher than that of men of the same age. In all of the paired population censuses, the relative question load is close to 1.

We restrict the data to individuals of eligible age and collapse it to the household-gender level. Then, we run the following regression specification for each DHS/MICS-census pair:

$$N_{ig} = \beta_0 + \beta_1 Female_g + \beta_2 MICS_i + \beta_3 (Female_g \times MICS_i) + \mu_{ig} \tag{2}$$

where N_{ig} stands for the number of household members of eligible age of gender g in household i, $MICS_i$ is an indicator for being interviewed by DHS/MICS (rather than the population census), and standard errors are clustered at the household level.

4 Results

4.1 Random assignment of man's questionnaire

We find that that the assignment of the man's questionnaire leads to the recording of a significantly lower number of men of eligible age in most surveys (see Figure 6). In 70% of the examined surveys, we estimate a statistically significant negative impact, implying a loss of men of eligible age between 3% and 21% across surveys.[3]

[2] We are currently adding 20 additional pairs to this analysis.

[3] We do not have power to identify effects smaller than 3%.

4.2 Difference-in-difference with population census

The difference-in-difference approach yields similar results for the woman's questionnaire. We detect a significant loss of women of eligible age in 8 out of 14 surveys, ranging between 2% and 10% (see Figure 8).

4.3 Age displacement vs. household member omission

By comparing how many missing household member of eligible age we detect to the number of additional household members of *ineligible* age recorded, we can decompose the loss of eligible members into two components: (i) age displacement - where enumerators manipulate respondents age to render them ineligible for individual questionnaires - and (ii) omission from household rosters - where enumerators do not record household members of eligible age at all.

Re-running the two main specifications for household members of ineligible age for the subset of surveys where we observe a statistically significant reduction in household members of eligible age, we find that in many surveys, excess men and women of ineligible age are recorded, respectively (see Figures 9 and 10). Dividing the absolute value of the absolute reduction in household members of eligible age by the absolute increase in members of ineligible age, we can determine the share of lost household members of eligible age whose age is displaced. We find that there is a lot variation across surveys in the share of members with a displaced age. In fact, in some surveys, the loss of members of eligible age is completely explained by age displacement while in other surveys it seems to be entirely driven by the omission of these members from rosters (see Figure 11).[4]

5 Selection out of sample

Who are the household members of eligible age that are screened out of individual questionnaires by enumerators? Answering this question is challenging because the missing household members are not directly observable, neither are their characteristics. But the comparison of recorded men of eligible age in households with and without the man's questionnaire is informative about the characteristics of the missing men. Differences in average characteristics between these two groups reflect selection of men out of sample.

[4]Note: that values above 1 and below zero are not statistically significantly different from these values.

Running specification (1) on individual-level characteristics recorded in the household roster (and thus observable for all men, independent from whether their households are eligible for the man's questionnaire or not), we find that missing men differ systematically from included men. In particular, men of eligible age recorded in households eligible for the man's questionnaire are older, more educated and more closely related to the household head in most surveys (see Figures 12, 13 and 14). This implies that missing men tend to be younger, less educated and less closely related to the head of their household.

6 Implications for aggregate statistics

Endogenous sample selection resulting from the screening of household members out of individual questionnaires has implications of aggregate statistics. First, it leads to a reduction in precision because less household members of eligible age are interviewed. Second, it leads to bias in aggregate statistics because samples are de facto not randomly selected, as shown in the previous section. We differentiate biases in two types of statistics that ensue, bias in individual-level statistics and bias in household-level statistics. We discuss each of them below.

6.1 Individual-level statistics

Bias in individual-level statistics arises because members that are screened out differ systematically from members that are not. Hence, the observed sample of men/women of eligible age is not representative of the population. Is the resulting bias quantitatively important? To address this question, we estimate national average educational attainment for men of eligible age in DHS/MICS sub-samples with and without the man's questionnaire. Figure 15 shows the difference between the two sub-samples. We find evidence of small differences (less than 5%)in most surveys. Only in a handful of surveys, national average educational attainment of men of eligible age is overestimated by more than 5% in the sub-sample eligible for the man's questionnaire.

While implications for average education at the national level seem to be limited, it is important to note that it would also be desirable to understand implications for statistics that are generated from data that is only collected in individual questionnaires, such as fertility statistics. Since fertility is not observed for men in the sub-sample of households that is not eligible for the man's

questionnaire, the presented approach does not allow us to assess this effect, however. Therefore, we are currently working on an alternative approach.

6.2 Household-level statistics

Household-level statistics will be biased in so far as they are normalized by household size and household size is biased due to the omission of household members from rosters. As figure 16 shows, household size is in fact significantly underestimated in a significant share of the examined surveys. One implication of this is that in many surveys, the national average number of household members per bedroom - a statistic that was of broad interest during the COVID-19 pandemic - is underestimated in households eligible for the man's questionnaire. This is likely only a lower bound of the total extent of underestimation, though, because the number of eligible female household members is likely to be underestimated throughout (since the woman's questionnaire is administered in all households).

7 Conclusion

Descriptive statistical analysis and causal inference lie at the core of empirical social science research. While causal inference was revolutionized by the introduction of experimental methods in the early 2000s and identification has been the subject of much methodological research since, data-generating processes have received considerably less attention. However, good data is paramount for both causal inference and descriptive analysis (Dillon et al. 2020).

This paper examines the production of household survey data, arguably one of the most important data sources in the social sciences. We show that enumerators systematically screen out household members that require disproportionate effort based on ex-ante observable characteristics (age and gender), either by omitting such household members from household rosters or by manipulating the eligibility criteria. This enumerator behavior induces selection of household members out of sample and as a result, aggregate statistics are biased. Preliminary evidence suggests that the magnitude of these biases varies significantly across contexts and domains.

Many open questions remain. Ongoing work of ours is focused on three aspects in particular. First, it aims to document the implications for aggregate statistics in more detail. Second, it

seeks to characterize the settings, households, survey respondents and enumerators particularly susceptible to endogenous sample selection in order to shed light on the underlying mechanisms. Third, it strives to develop remedies that help address endogenous sample selection ex-post through selection correction and ex-ante through survey design and implementation choices.

It is important to note that ex-ante observable variation in effort cost across household members is a universal feature of household surveys. Eligibility for individual questionnaires based on gender and age reported in the household roster is not only a standard component of DHS and MICS, it is also commonly observed in labor force, living conditions and household budget surveys. Thus, endogenous sample selection is unlikely to be limited to DHS and MICS. In fact, these two surveys are typically well-funded and implemented with external support, unlike many labor force and household budget surveys, and may thus suffer *less* from screening out of household members than other surveys

Figures

					Éligible pour :		
					QUESTIONNAIRE FEMME	MODULE TRAVAIL DES ENFANTS	QUESTIONNAIRE ENFANT DE MOINS DE 5 ANS
HL1.	HL2.	HL3.	HL4.	HL5.	HL6.	HL7.	HL8.
N° de ligne.	*Nom et prénoms*	QUEL EST LE LIEN DE PARENTE* DE *(nom)* AVEC LE CHEF DE MENAGE?	*(Nom)* EST-IL/ ELLE DE SEXE MASCULIN OU FEMININ ? 1 MAS. 2 FEM.	QUEL AGE A *(nom)*? QUEL AGE AVAIT *(nom)* A SON DERNIER ANNIVERSAIRE? *Enregistrer en années révolues* 98=NSP	*Encercler le numéro de ligne s'il s'agit d'une femme âgée de 15 -49 ans*	*Pour chaque enfant âgé de 5-14 ans :* QUI EST LA MERE OU LE PRINCIPAL GARDIEN DE CET ENFANT? *Enregistrer le n° de ligne de la mère/ gardien(ne)*	*Pour chaque enfant de moins de 5 ans :* QUI EST LA MERE OU LE PRINCIPAL GARDIEN DE CET ENFANT? *Enregistrer le no. de ligne de la mère/ gardien(ne)*
LIGNE	NOM ET PRENOMS	LIEN	M	F AGE	15-49	MERE	MERE
01		0 1	1	2 |___|___|	01	|___|___|	|___|___|
02		|___|___|	1	2 |___|___|	02	|___|___|	|___|___|
03		|___|___|	1	2 |___|___|	03	|___|___|	|___|___|

Figure 1: MICS, Togo 2006: Household roster

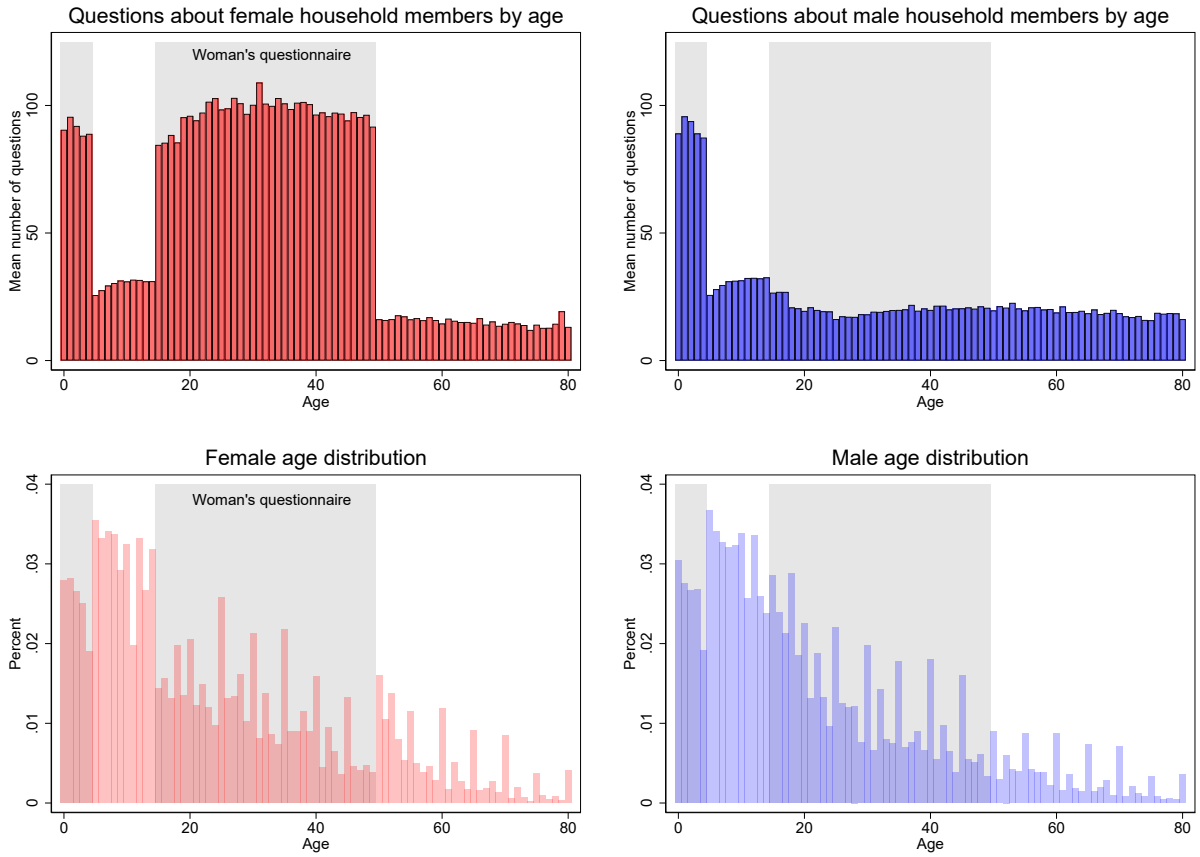

Figure 2: MICS, Togo 2006: Question load and age distribution by gender

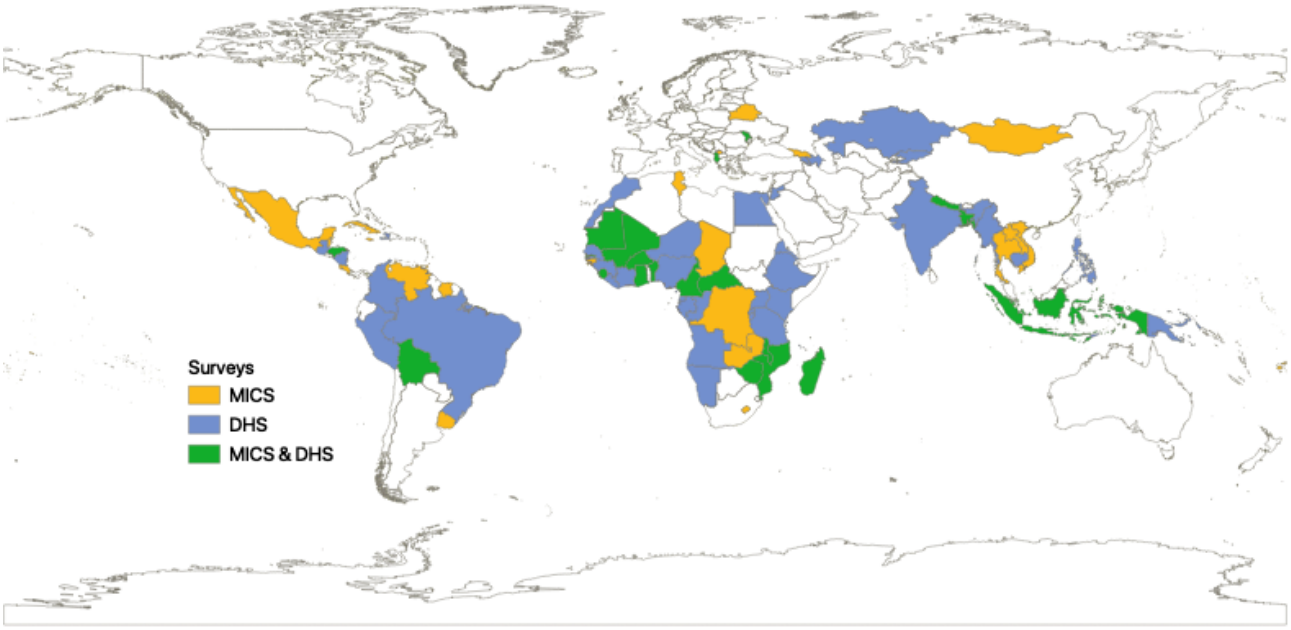

Figure 3: Geographic coverage of data

MICS

HOUSEHOLD QUESTIONNAIRE
GHANA 2011

HOUSEHOLD INFORMATION PANEL				HH
HH1. Locality Name Cluster No.: ___ ___ ___				HH2. Household Number: ___ ___
HH3. Interviewer name and number: ___ ___ ___				HH4. Supervisor name and number: ___ ___ ___
HH5. Date of interview: (DD/ MM / YYYY) ___ ___ / ___ ___ /2011				HH5A: Is the household selected for the male survey? Yes 1 No 2
HH6. Area: Urban 1 Rural 2	HH7.Region ___ ___	HH7A.District ___ ___	HH7B. Dist-type ___	HH7C. Sub-dist ___ ___
HH7D. Structure Address:				HH7E: Contact No of HH:

WE ARE FROM THE GHANA STATISTICAL SERVICE. WE ARE CONDUCTING A SURVEY THAT IS CONCERNED WITH FAMILY HEALTH AND EDUCATION. I WOULD LIKE TO ASK YOU A FEW QUESTIONS ON THESE AREAS. THE INTERVIEW WILL TAKE ABOUT 45 MINUTES. ALL THE INFORMATION WE OBTAIN WILL REMAIN STRICTLY CONFIDENTIAL AND YOUR ANSWERS WILL NEVER BE SHARED WITH ANYONE.

MAY I START NOW?
☐ Yes, permission is given ☒ Go to HH10 to get signature, then HH18 to record time, then begin interview.
☐ No, permission is not given ☒ Complete HH9. Discuss this result with your supervisor.

Figure 4: MICS, Ghana 2011: First page of household questionnaire

HH18. Record the time. Hour __ __ Minutes __ __

HOUSEHOLD LISTING FORM — HL

FIRST, PLEASE TELL ME THE NAME OF EACH PERSON IN YOUR HOUSEHOLD WHO USUALLY LIVES HERE, STARTING WITH THE HEAD OF THE HOUSEHOLD.
List the head of the household in line 01. List all household members (HL2), their relationship to the household head (HL3), and their sex (HL4)
Then ask: ARE THERE ANY OTHERS WHO LIVE HERE, EVEN IF THEY ARE NOT AT HOME NOW? (THESE MAY INCLUDE CHILDREN CURRENTLY IN SCHOOL OR AT WORK).
If yes, complete listing for questions HL2-HL4. Then, ask questions starting with HL5 for each person at a time.
Use an additional questionnaire if all rows in the household listing form have been used.

						For women age 15-49	For men age 15-59	For children age 5-14	For children under 5	For all household members	For children age 0-17 years			
HL1. Line number	HL2. Name	HL3. WHAT IS THE RELATION-SHIP OF (name) TO THE HEAD OF HOUSE-HOLD?	HL4. IS (name) MALE OR FEMALE? 1 Male 2 Female	HL5. WHAT IS (name)'S DATE OF BIRTH? 98 DK 9998 DK	HL6. HOW OLD IS (name)? Record in completed years. If age is 95 or above, record '95'	HL7. Circle line number if woman is age 15-49	HL7A. Check if HH5A=1 Circle line number if man is age 15-59	HL8. WHO IS THE MOTHER OR PRIMARY CARETAKER OF THIS CHILD? Record line number of mother/ caretaker	HL9. WHO IS THE MOTHER/ PRIMARY CARETAKER OF THIS CHILD? Record line number of mother/ caretaker	HL10. DID (name) STAY HERE LAST NIGHT? 1 Yes 2 No	HL11. IS (name)'S NATURAL MOTHER ALIVE? 1 Yes 2 No☒ HL13 8 DK☒ HL13	HL12. DOES (name)'S NATURAL MOTHER LIVE IN THIS HOUSEHOLD? Record line number of mother or 00 for "No"	HL13. IS (name)'S NATURAL FATHER ALIVE? 1 Yes 2 No☒ Next Line 8 DK☒ Next Line	HL14. DOES (name)'S NATURAL FATHER LIVE IN THIS HOUSEHOLD? Record line number of father or 00 for "No"
Line	Name	Relation*	M F	Month Year	Age	15-49	15-59	Mother	Mother	Y N	Y N DK	Mother	Y N DK	Father
01		01	1 2			01	01			1 2	1 2 8		1 2 8	
02			1 2			02	02			1 2	1 2 8		1 2 8	
03			1 2			03	03			1 2	1 2 8		1 2 8	
04			1 2			04	04			1 2	1 2 8		1 2 8	

Figure 5: MICS, Ghana 2011: Household roster

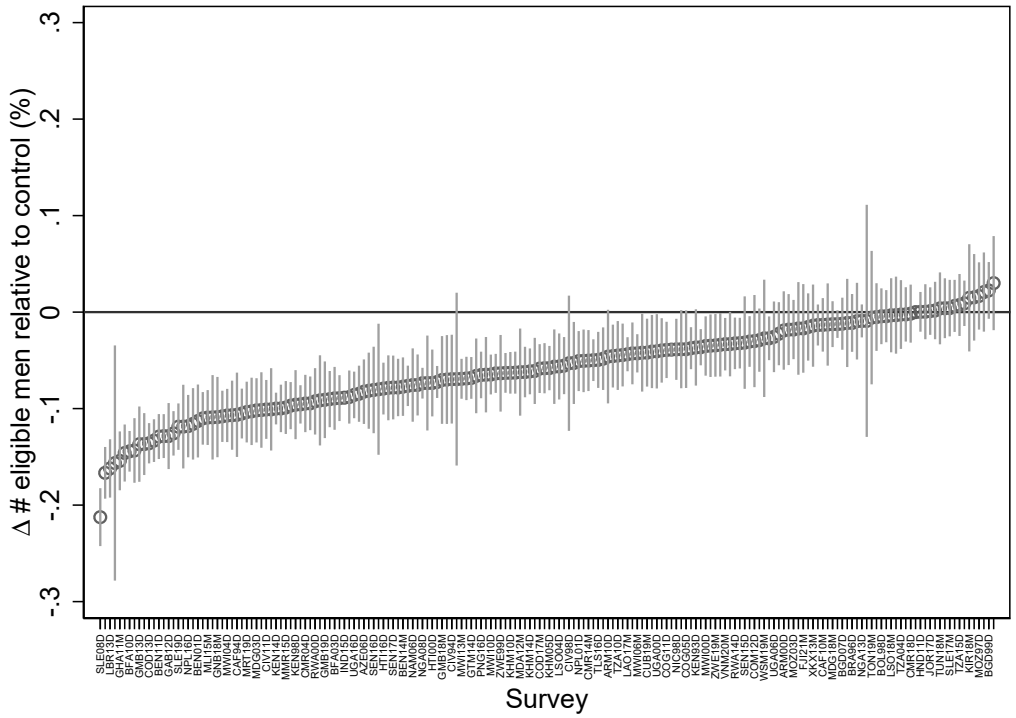

Figure 6: Missing men of eligible age in households eligible for man's questionnaire

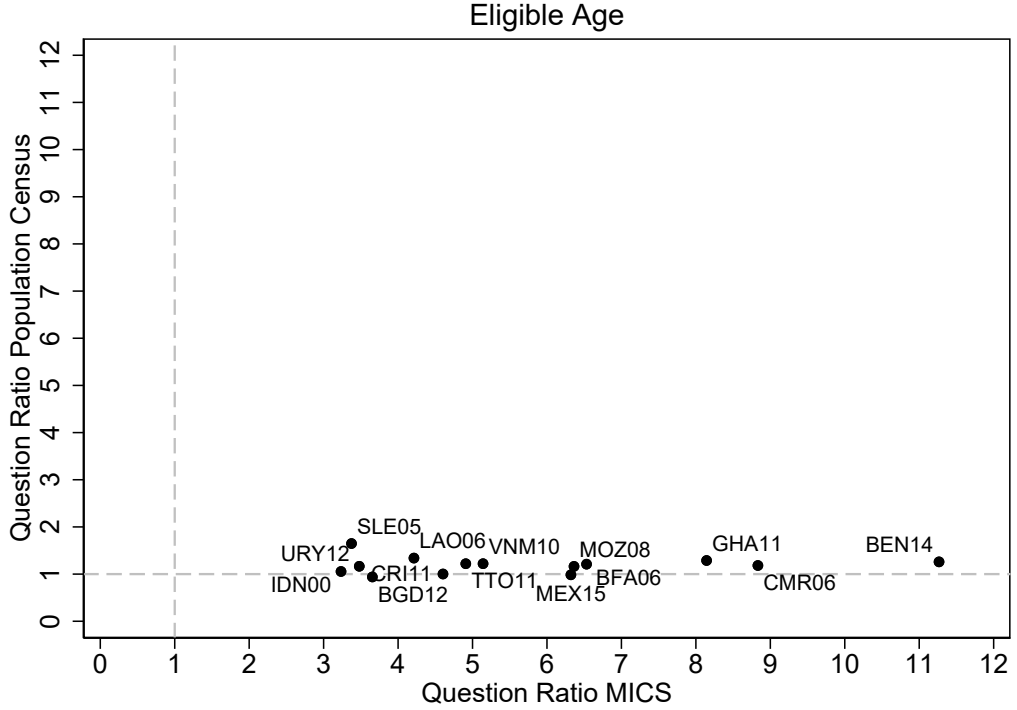

Figure 7: Relative question load of women of eligible age relative to men of the same age

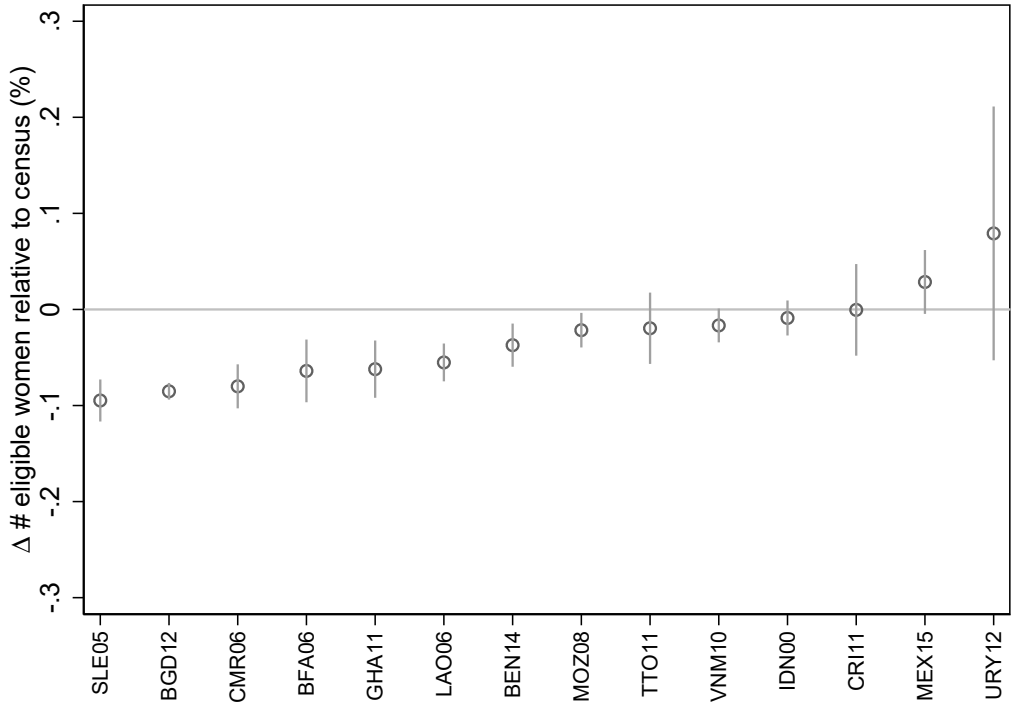

Figure 8: Missing women of eligible age in DHS/MICS relative to population census

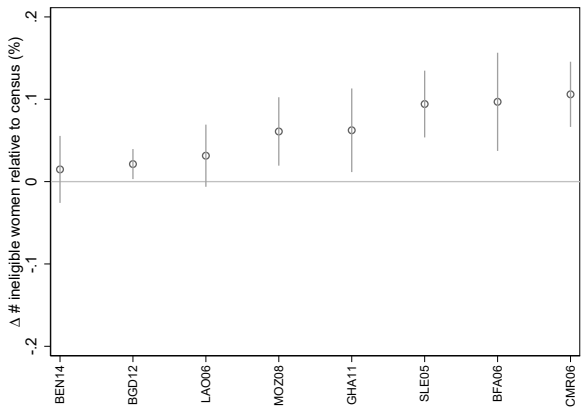

Figure 9: Excess women of ineligible age in DHS/MICS relative to population census

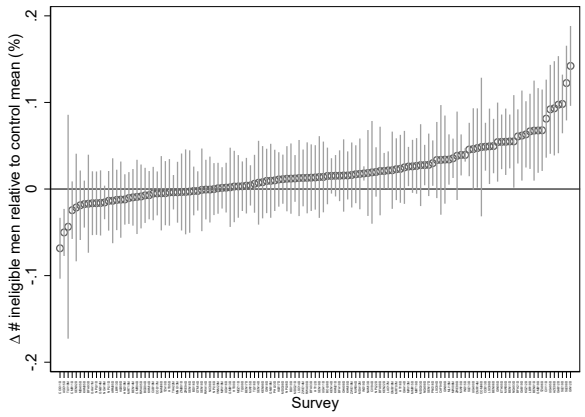

Figure 10: Excess men of ineligible age in households eligible for man's questionnaire

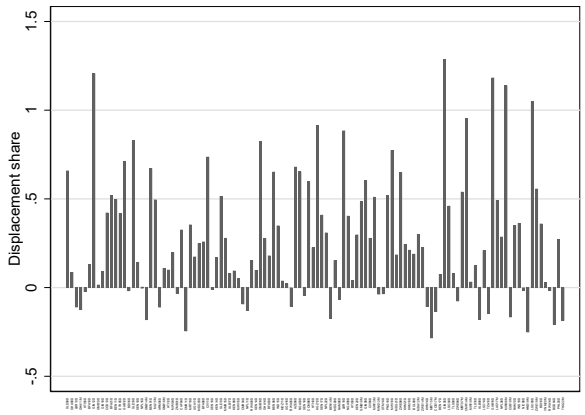

Figure 11: Share of missing men of eligible age with displaced age (RCT)

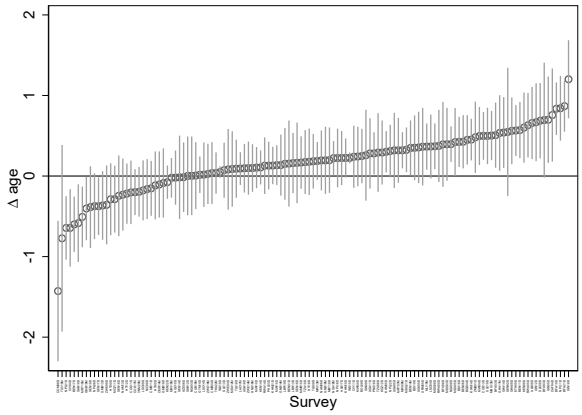

Figure 12: Effect of man's questionnaire on age of men in eligible age range

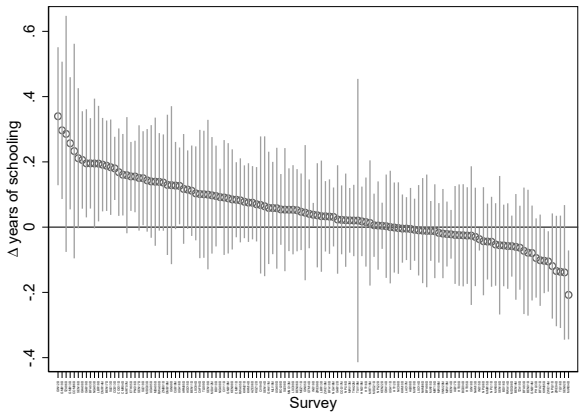

Figure 13: Effect of man's questionnaire on education of men of eligible age

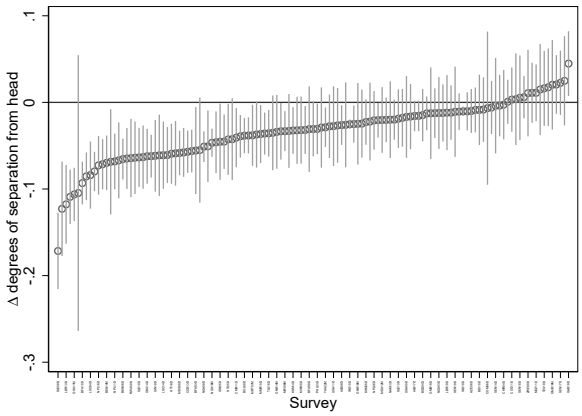

Figure 14: Effect of man's questionnaire on relationship to household head

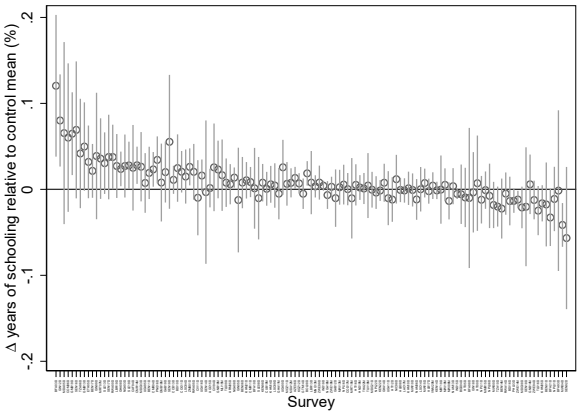

Figure 15: Effect of man's questionnaire on education of men of eligible age relative to control (weighted)

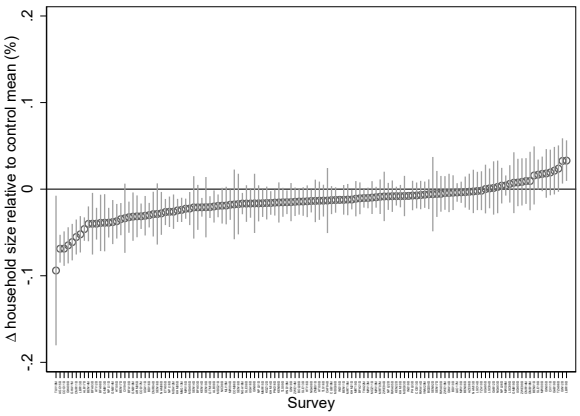

Figure 16: Effect of man's questionnaire on household size relative to control

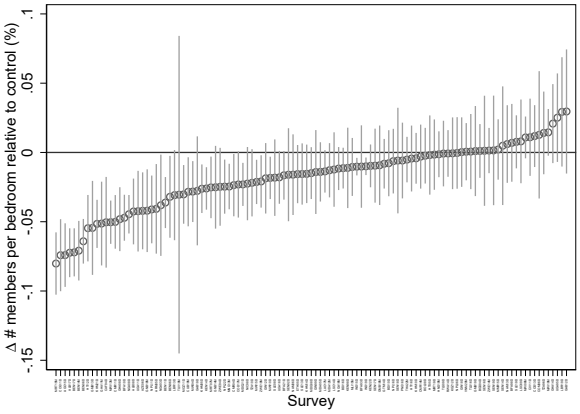

Figure 17: Effect of man's questionnaire on members per bedroom relative to control (weighted)

Tables

Table 1: MICS/DHS-Population Census pairs

Country	Survey	Survey Year	Census Year	Source of Census Data
BEN	MICS	2014	2013	IPUMS-International
BFA	MICS	2006	2006	IPUMS-International
BGD	MICS	2012	2011	IPUMS-International
CMR	MICS	2006	2005	IPUMS-International
CRI	MICS	2011	2011	IPUMS-International
IDN	MICS	2000	2000	IPUMS-International
LAO	MICS	2006	2005	IPUMS-International
MEX	MICS	2015	2015	IPUMS-International
MOZ	MICS	2008	2007	IPUMS-International
SLE	MICS	2005	2004	IPUMS-International
TTO	MICS	2011	2011	IPUMS-International
URY	MICS	2012	2011	IPUMS-International
VNM	MICS	2010	2009	IPUMS-International