Endogenous Sample Selection*

Torsten Figueiredo Walter[§] Niclas Moneke[¶]

March 12, 2025

This paper examines how incentives for data collectors shape the selection of sampling units. We provide causal evidence that data collectors respond to variation in effort cost across survey subjects by excluding high-cost subjects, thereby breaching protocol. Exploiting the random assignment of eligibility for individual interviews across 3.4 million households in 181 surveys worldwide, we find that in 110 (39) surveys at least 5% (10%) of eligible subjects are missing from the sample. Selection out of sample is systematic: missing subjects disproportionately come from marginalised populations. Using three applications, we illustrate how this selection undermines microeconomic and macroeconomic analysis alike.

Keywords: Data collection, moral hazard, sampling, selection
JEL classification: C81, C83, J22, O11, O12

*We thank Alison Andrew, Gharad Bryan, Raj Chetty, Andrew Dillon, Kevin Donovan, Dennis Egger, Luis Garicano, Maitreesh Ghatak, Doug Gollin, Sydney Gourlay, Stephane Helleringer, Lisa Ho, Talip Kilic, Gabriel Kreindler, Rohini Pande, Mark Rosenzweig, Martin Rotemberg, Nick Ryan, Chris Udry, Chris Woodruff and various seminar and conference audiences at BU, CSAE Annual, ESEM Rotterdam, IFS-UCL, IPA, NEUDC, Northwestern, Notre Dame, Nova, NYU, NYU Abu Dhabi, Oxford, PacDev, Philadelphia Fed, STEG Annual, UIUC, UVA, USC, WUSTL and Yale for helpful comments. Claudia Alcarra, Ana Barros, Frederic Cochinard, Muskaan Chopra, Ameni Hajji, Abigail Lamptey, Sophia Lin, Shifa Muhammad, Alfred Shauri, Jordan Simpson, Kristina Sisiakova, Kevishen Valeyatheepillay, Emma Wang and Ruike Zhang provided excellent research assistance. Financial support from STEG is gratefully acknowledged.

[§]New York University Abu Dhabi, Office A5-193, Abu Dhabi, UAE. Email: t.f.walter@nyu.edu
[¶]Department of Economics, University of Oxford, UK. Email: niclas.moneke@economics.ox.ac.uk

1 Introduction

What do the sensitivity of fertility to climate shocks in Sub-Saharan Africa, the missing middle in India, and the absent gender gap in employment in the National Longitudinal Survey of Youths 1997 have in common? All three phenomena feature in human-collected data, but are to an important degree artefacts of underlying data collection processes. In particular, these data collection processes all introduce incentives for data collectors to manipulate samples. The resulting *endogenous sample selection* introduces systematic bias, ultimately leading to erroneous inference.

Human-collected data are ubiquitous, ranging from surveys, censuses, medical records and police reports to laboratory test results. They are essential for research and policy alike. *In theory,* sampling for data collection is fully determined by protocol. *In practice,* data collectors exert considerable influence over sampling: they frequently have the means to alter the sample, and the incentive to do so. In particular, in settings where the actions of data collectors are imperfectly observable, they may exclude subjects that require disproportionate effort. These observations raise three questions. First, do data collectors systematically exclude subjects? Second, is the resulting selection of subjects out of the sample non-random? Third, does this selection affect inference and analysis, and if so, how?

How could data collectors manipulate samples, and why would they? Data collectors commonly execute two tasks: first, they screen for eligible units among a target population using pre-established eligibility criteria. This screening determines which units are included in the sample. Second, they collect detailed data about the sampled units. Hence, successful screening of units creates more work for data collectors, introducing an incentive to sabotage the screening, thereby manipulating the sample.

Consider a simple example for illustration. Household surveys, such as the Demographic and Health Survey (DHS) or the Multiple Indicator Cluster Survey (MICS), primarily collect information about children (aged 0-5) and women (aged 15-49) by administering long individual questionnaires about them. To identify eligible women and children in the first place, data collectors list all household members and screen for eligibility based on sex and age. This setup creates an incentive for data collectors to either manipulate members' eligibility criteria or omit eligible members entirely.

Figure 1 illustrates how these dynamics play out in the 2006 MICS from Togo: The top panels show that question load for eligible women and children of either sex is about three times as high as for ineligible household members. The bottom panels highlight how the associated age distributions lack mass in all age ranges that are eligible for individual questionnaires (grey-shaded areas) and have excess mass on the ineligible side of eligibility thresholds. Reassuringly, the male age distribution (bottom right) shows the same missing mass below the age of 5 as the female age distribution, but does not display missing mass between 15 and 49 (gold-shaded area), thereby suggesting a causal link between question load and sample inclusion.

This paper provides causal evidence from 181 surveys across 73 countries that variation in data collector effort cost across survey subjects leads to endogenous sample selection. We exploit the random assignment of question load across 3.4 million households to show that: first, data collectors manipulate survey samples by excluding high-cost subjects. Second, manipulation leads to non-random selection of subjects out of sample, resulting in under-representation of marginal populations. Third, selection introduces systematic bias in aggregate statistics and gives rise to erroneous inference. Fourth, endogenous sample selection is observed across many forms of human-collected data in high- and low-income countries alike.

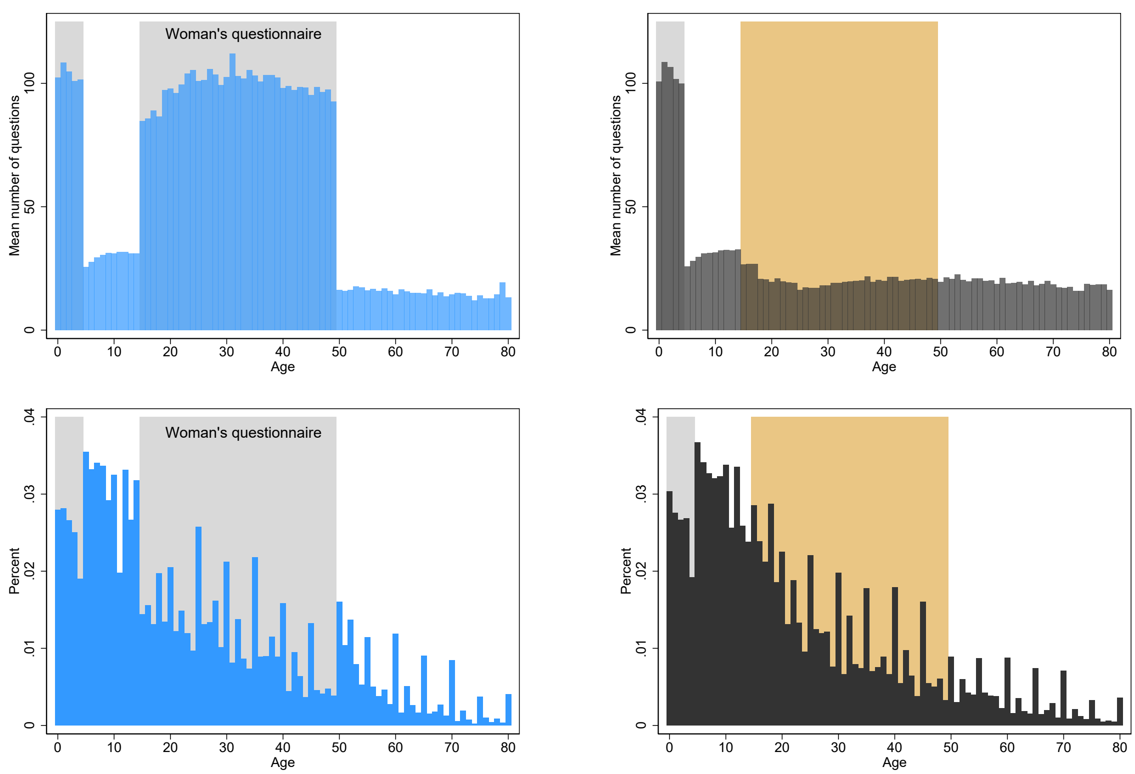


Figure 1: Question load (top) and age distribution (btm) by sex in Togo MICS 2006

Figure 1 plots the mean number of questions asked about female and male household members by age in panels (a) and (b), and their age distributions in panels (c) and (d). Age groups eligible for individual questionnaires are shaded in grey. Children under the age of 5 are eligible for the 'under-five questionnaire' and women between the ages of 15 and 49 are eligible for the 'woman's questionnaire'. In the right-hand panels, the age range between 15 and 49 is shaded in gold to facilitate comparison with the same range in the left-hand panels. Details on the measurement of question load by sex and age are available in Appendix A.1.3.

We view the data collector as an economic agent collecting data on behalf of their principal, for example a national statistical office or a team of researchers. The data collector's actions are imperfectly observable. In their decision to truthfully record a survey unit, e.g., a household member, as eligible or not for subsequent data collection, e.g., an individual questionnaire, the data collector weighs the effort cost of additional data collection against the expected penalty for manipulation. Thus, a simple empirical prediction is that survey units associated with higher effort cost and lower expected penalty are more likely to be excluded by data collectors.

To investigate endogenous sample selection empirically, we leverage data from two of the largest international household survey programs – the Demographic and Health Survey (DHS) and the Multiple Indicator Cluster Survey (MICS) – to document endogenous sample selection across a wide range of contexts, and to study its implications for aggregate statistics and economic analysis. The DHS and MICS are widely used in economics and beyond. Since 2010, at least 37 papers in top general interest economics journals have made use of these survey data, including Young (2012), Vogl (2013), Michalopoulos and Papaioannou (2016), Jayachandran and Pande (2017), Anderson (2018), Chatterjee and Vogl (2018), Hjort and Poulsen (2019), Corno et al. (2020), Lowes and Montero (2021), Guarnieri and Tur-Prats (2023), and Becker (2024).[1] Their use outside of economics is even greater.[2] At the same time, the DHS and MICS are of great importance for policy, e.g., the monitoring of the Sustainable Development Goals, aid allocation decisions, or to inform the design of national policies in low- and middle-income countries (Nolan et al., 2017).

To estimate the causal effect of data collector effort cost on sample exclusion, our main empirical strategy exploits the random assignment of individual questionnaires for men ('man's questionnaire') across households in 181 surveys (135 DHS and 46

[1] We include the American Economic Review, Econometrica, the Journal of Political Economy, the Quarterly Journal of Economics and the Review of Economic Studies in this count.

[2] In 2022 alone, close to 1,000 articles in reputable journals from the natural sciences, social sciences and fields of health and medicine referred to at least one survey program in their title or abstract. See Section 3 for details on the relevance of DHS and MICS in economics and beyond.

MICS). In this context, data collectors typically work on temporary contracts for the duration of each survey and receive a fixed daily wage. Extra individual questionnaires are time-consuming for data collectors, with each additional man's questionnaire requiring on average 25 minutes of extra work. Disutility of effort creates an incentive to shorten household interviews by reducing the number of household members eligible for individual questionnaires.[3] Indeed, we find that in 130 out of 181 surveys, the number of men eligible for the man's questionnaire is significantly smaller in households that have randomly been chosen to receive the man's questionnaire (henceforth referred to as treatment households). In the median survey, a randomly drawn man's questionnaire leads to a reduction in eligible men included in the survey by 6.5%. In a quarter of surveys, the reduction exceeds 9.3%.

Since the DHS and MICS were primarily designed to obtain information about women, and their children, in the population, the question remains if endogenous sample selection also affects eligible women – and to which degree given the considerably longer questionnaires for eligible women than for men in both survey programs. To address this question, we pursue a complementary empirical strategy to estimate the effect of the individual questionnaire for women ('woman's questionnaire'), which is key for deriving core survey outcomes such as fertility and child mortality. We cannot rely on random assignment of the woman's questionnaire for identification since such randomisation is extremely rare: as primary subjects of interest for the DHS and the MICS, women are almost always subject to long individual questionnaires.

Instead, we provide a complementary empirical strategy that compares the number of female household members of eligible and ineligible age in DHS and MICS to those in contemporaneous population censuses. Whereas in the DHS and the MICS the number of questions to be administered to women of eligible age (typically 15-49 years) is much larger than the number of questions to be administered to women outside this age range, no such difference in question load between women of eligible and ineligible age exists in population censuses. Therefore, whereas data collectors face strong incentives to exclude women of eligible age in DHS and MICS, no such incentive exists in census data collection. Based on 77 survey-census pairs with suitable micro data from 39 countries, we compare the average number of women of eligible and ineligible age in the household in the DHS/MICS to the average number of women in

[3] Additional time pressure for data collectors to finish the assigned number of household interviews per day further amplifies this incentive to exclude household members.

the census. We find strong evidence of endogenous sample selection closely mirroring our findings for men: we estimate that for 69 out of 77 survey-census pairs the lower bound of eligible women is significantly negative. In the median survey, the lower bound reduction in eligible women is 6.1%, exceeding 8.5% in a quarter of surveys.

The theoretical framework provides us with two testable empirical predictions. First, that exclusion of household members from the sample is increasing in data collector effort cost. Second, that it is decreasing in the probability of detection by their principal. We provide evidence in support of both predictions. Exploiting exogenous variation in day-to-day temperature within data collector, combined with the randomised assignment of the man's questionnaire, we can show the causal effect of higher effort cost from collecting data under extreme temperatures on the number of eligible men. We find an inversely U-shaped relationship, with temperatures above $70°F$, and especially above $80°F$, significantly increasing the number of eligible men missing in treatment households. In addition, we provide suggestive evidence comparing estimates of missing men from surveys that introduced systematic audits in the form of mandatory re-interviewing as an effective increase in the probability of detection. Indeed, the average share of missing eligible men amounts to 4.5% in surveys with audits compared to 7.3% in surveys without.

If women and men of eligible age were excluded from samples at random, our findings would be of minimal concern to data users. However, we show that missing individuals are non-randomly selected by data collectors and are best described as belonging to marginalised populations. By comparing eligible men in treatment and control households, we find that missing men are less closely related to the head of their respective household, younger, less educated, and less likely to have ever been married. A comparison of the characteristics of women of eligible age in the DHS/MICS and contemporaneous censuses yields the same conclusion for missing women. Reassuringly, in the few surveys without missing individuals, we fail to detect selection. For a subset of surveys, household rosters essential to investigate selection contain additional information: missing men also appear to be poorer, sicker and weakly more likely to be disabled. Taken together, our findings on selection suggest that among the high-effort-cost household members – men and women of prime age who are eligible for individual questionnaires – data collectors screen out exactly those individuals at the margins of their respective households, where household definitions and cultural norms leave room for interpretation, and chances that supervisors can detect manipulation are lower.

How does the relative under-representation of marginalised populations affect aggregate statistics? We show that two of the original core outcomes of the DHS and the MICS, fertility and marriage, are significantly upward-biased due to endogenous sample selection.[4] Put differently, the nuclear family, i.e., household members whose absence is easiest to detect, are relatively over-represented. Since these surveys remain a central data source on fertility for research (Vogl, 2016; Chatterjee & Vogl, 2018; Dupas et al., 2023) and an essential input for national health and education policies, the upward bias in aggregate is of concern in its own right.[5]

Does endogenous sample selection also affect inference and economic analysis, and if so, how? In three applications, we highlight that: First, endogenous sample selection can be correlated with a given treatment, in particular if the treatment affects the effort cost of data collectors. In such a case, even otherwise well-identified experimental or quasi-experimental estimates will be biased: we demonstrate that climate shocks in Sub-Saharan Africa are correlated with sample selection in the DHS, thereby leading to spurious relationships between shocks and outcomes of interest.

Second, to study labour and capital market frictions researchers often use firm or farm size distributions for inference, either explicitly by testing for bunching at policy-relevant thresholds or implicitly by leveraging moments of the firm size distribution to estimate model parameters. However, firm and farm censuses frequently employ size thresholds to determine sample inclusion and question load: we show that endogenous sample selection leads to distortions in recorded firm-size distributions. In particular, the entire 'missing middle' of firms in the Indian Economic Census can be explained by data collector incentives, with direct implications for structural estimation that uses these distorted firm-size distributions.

Third, endogenous sample selection can introduce initial selection of subjects into longitudinal surveys, generating dynamics in outcomes that are not representative of underlying population dynamics: we show that high-effort-cost individuals are missing from the US National Longitudinal Survey of Youths 1997 (NLSY97) and that the missing appear positively selected on family income. As a result, the dynamics of youth employment are diverging from those estimated from comparable survey data,

[4]Note that the predecessor of the DHS in the 1970s and 1980s was the World *Fertility* Survey.

[5]See Ministère de la Santé et de l'Hygiène Publique (2021) and Government of the Republic of Malawi (2018) for examples of the use of DHS fertility estimates for public policy and planning.

for example with respect to gaps across gender.

More fundamentally, our findings highlight a novel information-bias trade-off in data collection. As the number of questions to be administered increases, samples shrink. This, in turn, induces selection and leads to bias. To quantify this trade-off, we estimate the elasticity of sample size with respect to question load using both of our empirical strategies. We find an average elasticity of approximately -0.01, suggesting that adding an individual questionnaire that includes the same number of questions as the household roster to a survey leads to a reduction in eligible survey subjects by 1%. This elasticity remained remarkably stable over the history of the DHS and MICS, while question load proliferated dramatically, such that average missing men increased from 6.1% in the 1990s to 8.9% today.

How can endogenous sample selection in existing data be corrected for, and prevented in future data collection? Exploiting either contemporaneous censuses or control group survey households to provide information about marginal distributions of population parameters enables re-weighting, such as iterative raking. For example, we show how multivariate raking can reduce approximately half of the estimated bias in fertility. In other words, data collectors appear to select based on a combination of observable and unobservable (to the econometrician) respondent characteristics.

For *ex ante* prevention, we investigate the role of technology and recent innovations in the DHS and MICS how sample selection could be reined in. We find suggestive evidence that tablets and field check tables alone are ineffective, whereas mandatory re-interviewing markedly reduces the number of missing individuals. Alternatively, to limit the effects of incentives, a simple policy recommendation is to ensure the division of labour between listing and data collection teams. Of note are also the large benefits of introducing randomised elements in data collection designs – both for live diagnostics and *ex post* correction.

Finally, beyond our main analysis, we assemble additional suggestive evidence to confirm how widespread of an empirical phenomenon endogenous sample selection as a result of data collector incentives appears to be – affecting numerous forms of data collection and popular data products in low- and high-income countries alike.

This paper contributes to four streams of literature. First, it adds to a long and active literature on selection in surveys (Rubin, 1976; Meyer et al., 2015; Dutz et al., 2021). While this literature is largely focused on non-response bias, i.e., self-selection of respondents, this paper highlights an overlooked margin of selection that, unlike non-response, is not (directly) observable to the econometrician: the selection of respondents by data collectors. We show that endogenous sample selection driven by data collector incentives can lead to substantial bias in aggregate statistics and erroneous inference.

Second, this paper contributes to a broad literature on the design and implementation of data collection. While the manipulation of respondent screening by data collectors has been recognized as a potential concern by practitioners (Marckwardt & Rutstein, 1996; Pullum, 2006), it has barely received any attention from researchers.[6] We examine the screening incentives of data collectors through the lens of economic theory and provide systematic, global evidence of their impact on sample selection. In doing so, we relate to at least three streams of work within the broader literature on data collection. First, we relate to research on enumerator effects (West & Blom, 2017; Di Maio & Fiala, 2019). While we also examine the role of data collectors in data collection, we do not focus on the effect of the identity of data collectors on data collection, but investigate how variation in incentives (within data collector) affects data collection. Second, we relate to recent work on respondent fatigue (Ambler et al., 2021; Abay et al., 2022; Jeong et al., 2023) examining the effect of question load on respondent behaviour. In this paper, we instead shed light on the effect of question load on the behaviour of data collectors. Finally, we contribute to the wider literature on survey design – including research on question design (Bardasi et al., 2011; Dillon et al., 2012; Serneels et al., 2017), respondent effects (Kilic et al., 2021; Dervisevic & Goldstein, 2023; Dillon & Mensah, 2024; Masselus & Fiala, 2024), interviewer pay (Menold et al., 2018) and supervision arrangements (Sen, 2024). We provide new insights on how the allocation of question load across sampling units translates into incentives for surveyors to falsify survey responses with regards to eligibility criteria.[7]

Third, we provide a cautionary tale for the study of climate shocks. We demonstrate that extreme weather events, such as heat, droughts and floods, affect sample composition in data collected by humans on the ground. This endogenous sample selection, in turn, undermines causal identification of the impact of climate shocks on any outcomes from such data – with implications for a wide range of work investigat-

[6] An exception is Eckman and Koch (2019) who compare the European Social Survey, in which interviewer involvement in sampling varies across countries, with adjacent labour force surveys.

[7] This paper is also related to the cross-disciplinary literature concerned with data fabrication in surveys and its implications for research (Crespi, 1945; Kosyakova et al., 2014; Blasius & Thiessen, 2015; Finn & Ranchhod, 2015; Robbins et al., 2018; Castorena et al., 2023).

ing the effects of climate shocks on health (Burke et al., 2015; Fichera & Savage, 2015; Nagata et al., 2021; Le & Nguyen, 2022), fertility (Norling, 2022), mortality (Geruso & Spears, 2018), marriage (Corno et al., 2020; Corno & Voena, 2023), domestic violence (Epstein et al., 2020), consumption (Paxson, 1992; Dimitrova, 2021), wealth (Thiede, 2014) and conflict (Miguel et al., 2004; Couttenier & Soubeyran, 2014).

Fourth, we contribute to the literature on the missing middle. Whether or not firm size distributions in low- and middle-income countries exhibit a relative lack of medium-sized firms remains subject of ongoing debate (Tybout, 2000, 2014; Hsieh & Olken, 2014; Abreha et al., 2022). We propose endogenous sample selection as a novel explanation for the observation of a missing middle. In particular, we argue that design and implementation of firm censuses frequently generate incentives for data collectors to omit medium-sized firms, as evident in the Indian Economics Census.

The remainder of the paper is structured as follows. Section 2 introduces a simple theoretical framework. Section 3 provides background on the relevance, design and implementation of the DHS and MICS. In Section 4, we present empirical strategies and results on missing individuals. In Section 5, we examine the selection of missing individuals. Section 6 presents three applications of endogenous sample selection highlighting how empirical analysis can be biased as a result. Finally, Section 7 discusses the relevance of endogenous sample selection for data collection more broadly, before Section 8 concludes.

2 Theoretical framework

A simple theoretical framework of data collector behaviour can guide the empirical analysis of endogenous sample selection. Data collectors face a choice of including or excluding a given subject from the eligible population in the initial listing exercise.[8]

The decision to report or not to report a given subject, $R \in \{0, 1\}$, in the listing is determined by two competing forces: the cost c of enumerating a given subject if the data collector reports them, and the probability of detection p (and associated penalty of losing their wage w) if they do not report. Both forces vary with the subject's observable characteristics x, such as a their age or sex.

[8] This decision to exclude covers both empirically relevant cases we document in Section 4: displacement, i.e. adjusting a subject's age to make them ineligible, and omission, i.e. not reporting a subject on the roster. Both cases lead to exclusion of that subject from the long questionnaire.

The utility of the data collector, U, as a function of the decision to report, R, can be written as follows:

$$U(R) = (1 - [p(x) \times (1 - R)])w - c(x) \times R \quad (1)$$

The data collector will decide to report a given subject $R = 1$ if their wage w exceeds the detection probability-scaled cost of enumeration:

$$U(1) > U(0) \quad if \quad w > \frac{c(x)}{p(x)} \quad (2)$$

Therefore, a given subject is more likely to not be reported if their enumeration cost is high, or the detection probability of them not having been reported is low. As illustrated in Figure 1 for household surveys with individual-level questionnaires, the cost of enumeration, driven by question load, varies across household members as a function of their observable characteristics, vector x, e.g., age and sex. For example, a 16-year old female is eligible for a long questionnaire in the DHS/MICS surveys, incurring higher cost for data collectors than a 14-year old female or 16-year old male.

However, deciding not to include a subject entails a risk of detection by the supervisor. With probability p manipulation in the form of not reporting a subject is detected, the data collector gets fired and does not receive their wage w. Detection is not random, and we assume the detection probability to depend on observable characteristics x. In our setting, some observables that aid detection also drive data collector cost, such as a household member's age, whereas other observables are independent of cost: the genealogical distance to the household head, ceteris paribus, does not affect data collector cost – but a household without a wife or husband is much easier to detect for supervisors than one missing an aunt, cousin, niece or grandchild.

The main testable prediction arising from our theoretical framework is that data collectors maximise utility of data collection by not reporting, i.e. excluding, high cost, low detection probability subjects. In Section 4 we test if data collectors exclude household members from the sample in general. Exogenous variation in data collector effort cost (or variation in detection probability) allows us to explicitly test the mechanism described above (cf. Subsection 4.3). Finally, if data collectors exclude household members as a function of subjects' expected effort cost and the perceived probability of detection for a given subject, this can introduce systematic selection of the sample which we test for in Section 5.

3 Background

3.1 Relevance

In this paper, we primarily 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, HIV/AIDS, malaria, and nutrition. It is funded by USAID and implemented by ICF. Since its inception in 1984, the program has conducted more than 400 surveys, with sample sizes ranging between 5,000 and 12,000 households and an estimated average cost of USD 1.6 million per survey. The MICS program bears many similarities with the DHS. It also mainly focuses on the situation of children and women in low- and middle-income countries and comprises more than 350 surveys. Sample sizes and costs tend to be lower, however, averaging around 12,000 households and USD 1.1 million per survey, respectively. Both survey programs have a reputation for collecting accurate, comparable, nationally representative data using standardized, state-of-the-art survey instruments across countries.[9]

We focus on these two household survey programs for three reasons. First, they are of great importance for research, especially in the social sciences and the fields of medicine and health.[10] As data from the Web of Science database shows (Figure A1), the annual number of articles published in reputable journals across all fields that refer to the DHS and the MICS in their title or abstract has increased 27-fold since 2000, reaching nearly 1,000 in 2022.[11] The true use of the data is likely much higher, though, because many papers that use the data do not explicitly mention them in title or abstract. For example, out of at least 37 papers published in Economics top 5 journals since 2010 that use the DHS or the MICS, only one refers to them in title or abstract.

Second, the DHS and the MICS are of great importance for policy. They are key to monitoring the Sustainable Development Goals (SDGs), providing input data for about 30 SDG indicators. They affect aid flows, not least through programs that are explicitly conditioned on DHS-derived indicators, such as the World Bank

[9] See Sustainable Development Solutions Network (2015) for details on survey cost estimates.

[10] Short Fabic et al. (2012) provide a historic overview of DHS use in population/health research.

[11] Statistics based on Web of Science database keyword search, restricted to journals that formed part of the Essential Science Indicator journal master list as of June 2024.

Program-for-Results. At the national level, they are an important input to policy, in particular in health sector, as documented by Nolan et al. (2017) and evidenced by frequent references to them in national health policy plans (Ministry of Health, Republic of Ghana, 2020; Ministry of Health, Republic of Kenya, 2022; Government of the Republic of Malawi, 2018; Ministry of Health, Uganda, 2017).

Third, both survey programs are of *global* importance. Since program inception, the DHS and the MICS program have conducted surveys in more than 90 and 120 countries, respectively, making them a unique source of globally comparable data over a time span of more than 30 years.[12]

3.2 Survey design

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 questionnaire. 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 sex 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 phases, both survey programs introduced a man's questionnaire. This questionnaire addresses similar topics as the woman's questionnaire – mainly fertility, health and sexual behaviour – but is typically much shorter. In most surveys, the eligible age ranges from 15 to 49, but in some cases it also includes older men up to the age of 54, 59 or 64. Importantly, in many surveys this questionnaire is only administered in a random subset of households within each enumeration area.

Individual questionnaires are administered after the household roster has been

[12]Statistics retrieved from the official DHS website – https://dhsprogram.com/ – and the official MICS website – https://mics.unicef.org/ – on August 18, 2024.

completed. This implies that at the time of the roster completion, survey respondents do not know how the age and sex of household members recorded in the roster affect the length of the household interview. Data collectors 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 data collectors to mark every eligible member as they fill in the roster (see Figure A3 for illustration).

An important difference between the DHS and the MICS lies in the household definition they work with. The MICS operates with a de jure household definition, recording all usual members. Each of these members qualifies for the individual questionnaire if they are in the eligible age range. The DHS instead records all usual household members *and* all guests who stayed in the household the night before.

The eligibility of de facto and de jure household members for individual questionnaires, however, varies across surveys. In phases 1 and 2 of the DHS program, eligibility was conditional on having slept in the household last night. From phase 3 onwards, most surveys did not condition eligibility on having slept in the household last night anymore. However, all results published by the DHS remain restricted to de facto populations to avoid double-counting.[13] Therefore, we define eligibility in the DHS as being of eligible age and having slept in the household last night throughout our analysis.

3.3 Data collector incentives

DHS and MICS are funded and supported by USAID and UNICEF, respectively. Both programs provide questionnaire templates that are standardized within survey phases and guidelines for implementation in the form of manuals for data collectors, supervisors, editors as well as data collector training, household sampling and other topics. However, surveys are ultimately implemented by local agencies, most commonly National Statistical Offices.[14] Hence, data collectors are recruited locally. Nonetheless, hiring practices barely vary across contexts. Temporary contracts for the duration of the survey are standard. Only a few implementing agencies rely on

[13] In fact, none of the data from individual interviews of household members who did not sleep in the household last night is published.

[14] 82% of the surveys in our main sample were implemented by National Statistical Offices, 15% by other governmental bodies, such as Ministries of Health, and 3% by nongovernmental organizations.

their permanent staff for enumeration in addition to temporary workers.[15] Data collectors generally have to meet the following criteria: They have to (i) be available to work full time for the duration of the survey, (ii) exceed a minimum level of physical fitness, so they can walk long distances, and (iii) speak at least one of the languages used for training. Additionally, there is a preference for local candidates from within a region of a country and candidates with secondary or higher education. As a result, interviewers are more educated than the average respondent in most contexts.

Data are collected by enumeration teams usually comprised of a supervisor, a field editor and several data collectors. Supervisors are in charge of the organization of the fieldwork, including the assignment of households and questionnaires to data collectors and spot check re-interviews. Field editors are responsible for monitoring data quality. To this end, they observe interviews, edit completed questionnaires and may ask data collectors to return to interviewed households to correct problems. Additional data quality issues can be detected through field check tables produced by data processing teams during fieldwork. These are typically provided to supervisors after the completion of an enumeration area and can inform measures to improve data quality going forward. All of this implies that the missing eligible individuals we detect in this paper were either not flagged in any of the data quality checks or, if flagged, they were not addressed successfully.[16]

Data collectors' employment contracts are designed by the implementing agencies. Thus, they can vary across surveys. In practice, however, data collectors are almost always paid a fixed daily wage plus a per diem for food and accommodation. The daily workload of enumeration teams is typically set in advance by the central office of the implementing agency and adherence to the schedule is heavily emphasized during fieldwork. Supervisors are responsible for assigning households to data collectors at the beginning of each day, but these assignments can be adjusted throughout the day as some interviews take shorter or longer than expected. Data collector performance is monitored continuously throughout the survey. Supervisors complete a so-called 'interviewer progress sheet' after the completion of each survey cluster to track how

[15] Fieldworker data from recent DHS confirm that most data collectors work under temporary contracts. In the 19 surveys included in our main sample for which fieldworker data is available, on average 13% of data collectors are permanent employees and 87% have temporary contracts.

[16] Neither in the DHS nor the MICS data is it possible to observe which interviews were monitored by a field editor or re-conducted by a supervisor.

data collectors are keeping up with the assigned workloads.[17] This means that data collectors benefit from missing eligible household members in at least two ways. First, they will be better able to keep up with the assigned workloads, thereby building a good reputation, minimising their risk of termination, and increasing their chances of re-employment.[18] Second, they may have shorter working days.

The incorrect completion of household rosters also carries a risk for data collectors. Supervisor guidelines indicate that terminations may be necessary in cases of data falsification. It is unclear how common such terminations are in practice, but the DHS recommends implementing agencies to recruit reserve data collectors who can step in after separations.[19]

4 Missing household members

4.1 Missing men

4.1.1 Empirical strategy

In many DHS and MICS, only a random subset of households is eligible for the man's questionnaire. Random assignment is carried out at the headquarters of the implementing agency after household listing in all enumeration areas has been completed.[20] To this end, USAID and UNICEF provide implementing partners with a computer tool for randomisation.[21] Each household's randomly drawn eligibility status is prefilled on their questionnaire and thus visible to data collectors.

Relying on the random assignment of the man's questionnaire across households, we run the following OLS regression to estimate the causal effect of eligibility for the man's questionnaire:

$$Y_{ic} = \alpha_c + \beta MQ_{ic} + \epsilon_{ic} \tag{3}$$

where Y_{ic} is an outcome of interest of household i in stratum c. MQ_{ic} is an indicator variable that takes the value one if household i is eligible for the man's questionnaire, and zero otherwise. α_c is a set of stratum fixed effects. In most surveys, strata

[17] See LoPalo's (2023) Online Appendix Figure 1 for the DHS 'interviewer progress sheet'.

[18] DHS fieldworker data shows that many data collectors have previous DHS experience.

[19] This subsection is based on exchanges with UNICEF's Data Collection Unit, and LoPalo (2023).

[20] The listing of households in all enumeration areas selected for a survey is typically carried out a few month ahead of the planned survey fieldwork.

[21] The MICS sampling and randomisation tool is available here.

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, as recorded during the household listing exercise preceding the survey. The regression coefficient β captures the causal effect of household assignment to the man's questionnaire on the outcome of interest.[22]

We attribute the difference in outcomes between treatment and control households to the difference in incentives faced by the data collector. While we concede that the listing of household members is ultimately the product of the interaction between the data collector and the respondent – and the respondent may also be interested in shortening the interview – we argue that the respondent does not possess the necessary information that would allow them to intentionally reduce the number of eligible men. First, respondents are unlikely to be familiar with the question load distribution across age since neither the DHS nor the MICS are panel surveys, but repeated cross-sectional surveys that are typically only carried out every 5 years.[23] Second, unlike data collectors, respondents do not know the eligibility of their household for the man's questionnaire.

4.1.2 Data

Based on the universe of survey reports published on the official DHS and MICS websites, we identify 181 surveys, 135 DHS and 46 MICS, carried out across 73 countries between 1991 and 2022 in which a man's questionnaire was administered to a random subset of households. Table A2 provides a complete list of these and Figure 2 illustrates their geographic coverage, including low- and middle-income countries from all continents. The resulting dataset includes 3.4 million households out of which 1.1 million were randomly assigned a man's questionnaire.[24]

The random assignment of the man's questionnaire to households is stratified by enumeration area. The treatment probability varies between 1/12 and 2/3 across

[22] Note that we do not observe any cases of eligible men in eligible households that were not attempted to be interviewed individually. Hence, data collectors appear to comply perfectly with the random assignment. But not all eligible men in eligible households complete the individual interview. Some are absent, incapacitated or refuse. The average completion rate in our sample of surveys is 90.5%.

[23] The continuous DHS in Senegal and Peru are exceptions to this. They are carried out annually.

[24] We identify additional surveys with a man's questionnaire that is randomly assigned across households. We do not include these here because either their design differs in important ways from the one described in Section 3.2 or the available microdata does not lend itself to our analysis. Details are provided in Appendix A.1.1. We also exclude surveys that do not have national coverage.

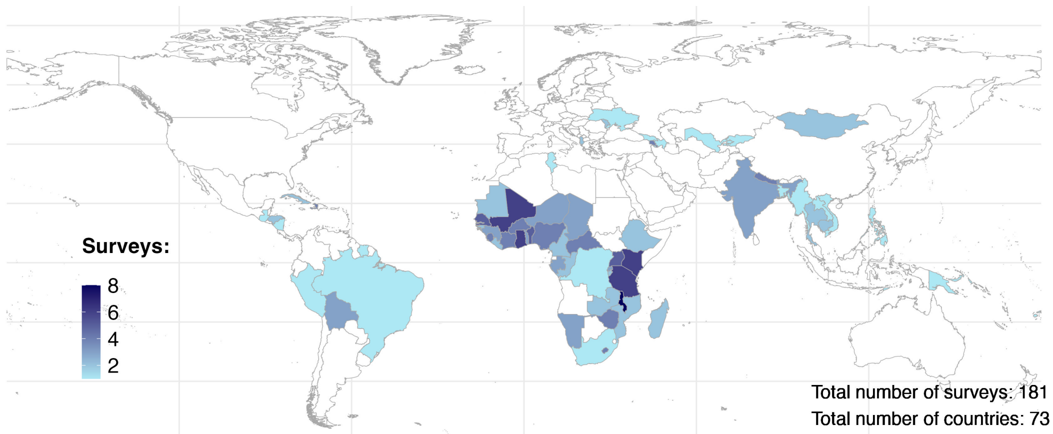


Figure 2: Geographic coverage of surveys with randomly assigned man's questionnaire

surveys, but it is most frequently 1/2 (in 55% of surveys) or 1/3 (in 34% of surveys). The median duration of the man's questionnaire varies between 6 and 50 minutes across surveys, with the average man's questionnaire lasting 25 minutes.

In a subset of surveys (76), men and/or women in treatment households who are eligible for the individual questionnaire as well as children under the age of 5 are also eligible for biomarker collection. This typically amounts to a combination of HIV testing among eligible adults, anaemia testing among eligible women and children, and malaria testing and anthropometry among children. Men's biomarkers are collected in 58 of these surveys. In all of these cases, we estimate the joint impact of the man's questionnaire and biomarker collection.

Microdata for the identified surveys is obtained from the DHS (ICF, 1982-2022) and MICS (UNICEF, 2000-2022) online microdata archives. All variables required for the analysis are harmonised across datasets, as detailed in Appendix Section A.1.4.

4.1.3 Results

We find that household assignment to the man's questionnaire leads to the recording of a significantly lower number of eligible men in most surveys. Figure 3 plots the point estimates and 95% confidence intervals of the β coefficient from specification (3) relative to the control mean, sorted by magnitude across surveys. We estimate a statistically significantly negative impact in 130 out of 181 surveys (72%). For the remaining 51 surveys, our point estimates are mostly negative, but insignificant (36 surveys). Only for a single survey, we estimate a statistically significant positive effect. The median reduction in eligible men amounts to 6.5%. In 25% of surveys the reduction exceeds 9.3%, peaking at 23%.[25]

Surveys with longer man's questionnaires display more missing men. As shown in Figure A6, an increase in the length of the man's questionnaire by 69 questions, corresponding to the difference between the 25th and the 75th percentile in the distribution of questionnaire length in our sample of surveys, is associated with an increase in missing eligible men by 2.2 percentage points.

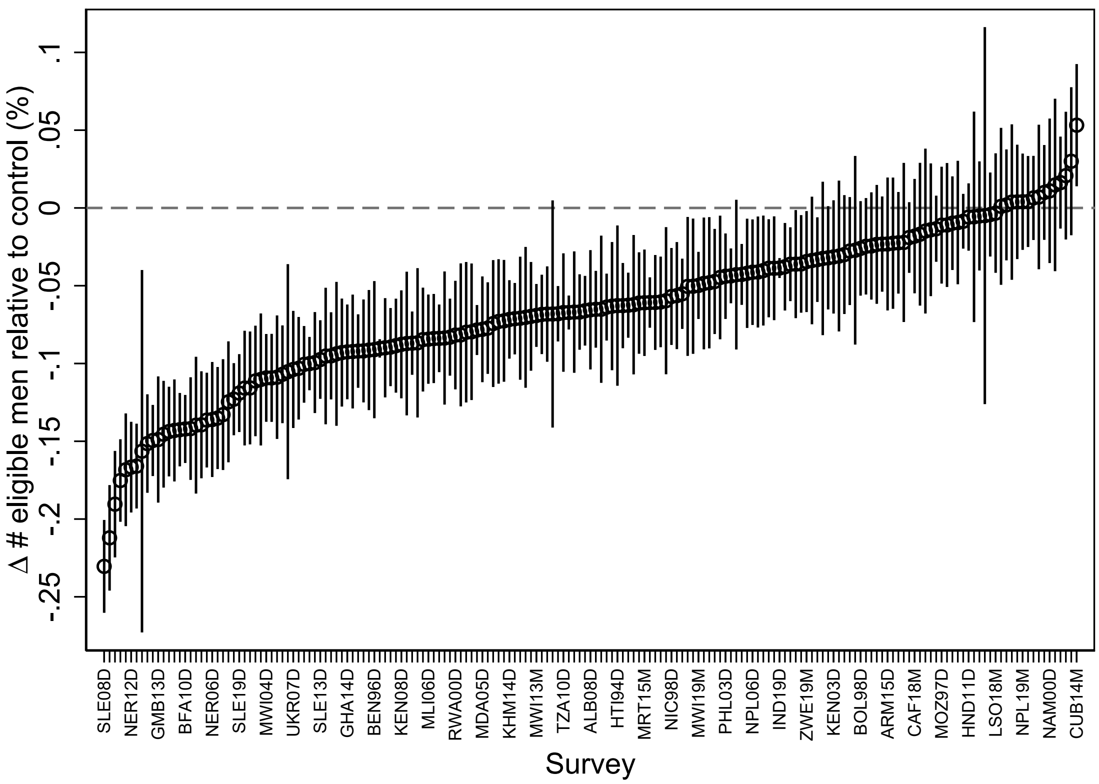


Figure 3: Effect of man's questionnaire on number of eligible men in the household

This figure displays estimates of β from equation (3) relative to the control mean where the outcome variable is the number of eligible men in the household. The sample consists of all 181 DHS and MICS with a man's questionnaire that is randomly assigned across households. Circles indicate point estimates and bars indicate 95% confidence intervals. Surveys are sorted along the x-axis in ascending order of the point estimate. Every 5th survey is labelled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively. All estimates are reported in Table A4, column (3).

Data collectors can achieve the observed reductions in the number of eligible

[25] Note that effects are larger in surveys where male biomarkers are collected alongside the questionnaire (see Figure A5).

men in treatment households in at least three ways. First, they can manipulate the eligibility criteria of household members such that they do not qualify for the man's questionnaire. Second, they can omit eligible men entirely from the household roster. Third, they can behave in ways that reduce the household response rate among eligible households with a large number of eligible men.

We find strong evidence in support of the first two margins, manipulation of eligibility criteria and omission of eligible men, but not the third. The number of ineligible men in treatment households exceeds the one in control households in many surveys, consistent with manipulation of eligibility criteria.[26] As shown in Figure 4, our point estimates are significantly positive for 56 surveys, significantly negative for 6 surveys and statistically insignificant in the remaining 119 surveys. Reassuringly, the total number of men in households is not affected positively by treatment in any surveys. Instead, it is either unaffected by treatment (94 surveys) or negatively impacted (87 surveys). The latter indicates that in many surveys, omission of eligible men from rosters is an important channel through which data collectors reduce their workload (see Figure A9).

We find limited evidence of differential household response by household assignment to the man's questionnaire. First, note that household response rates in the surveys under study are very high. In fact, the average survey in our sample has a response rate of 97.4%.[27] Hence, there is limited scope for differential response. Second, MICS data allow us to explicitly test for balance in response.[28]. We find that response is balanced between treatment and control in all but 5 out of 43 surveys (see Figure A4). In all of these five cases, treatment is associated with marginally lower response rates, with the shortfall ranging between 0.3 and 1.4 percentage points. Hence, strategic manipulation of household responses does not appear to be an important margin of data collector response.

It is worth noting that the administration of the man's questionnaire goes hand in hand with a change in data collectors. A strong emphasis on same-sex interviews in the DHS and the MICS program means that a male interviewer is required for

[26] Ineligible men are those who do not qualify for the man's questionnaire because of their age. In the DHS, ineligible men additionally include those who did not sleep in the household last night, independent of their age.

[27] Response rates are sourced from the final reports of all surveys in our sample.

[28] The DHS program does not publish data on households that did not complete the interview. However, the MICS program provides this data for all but three surveys in our sample.

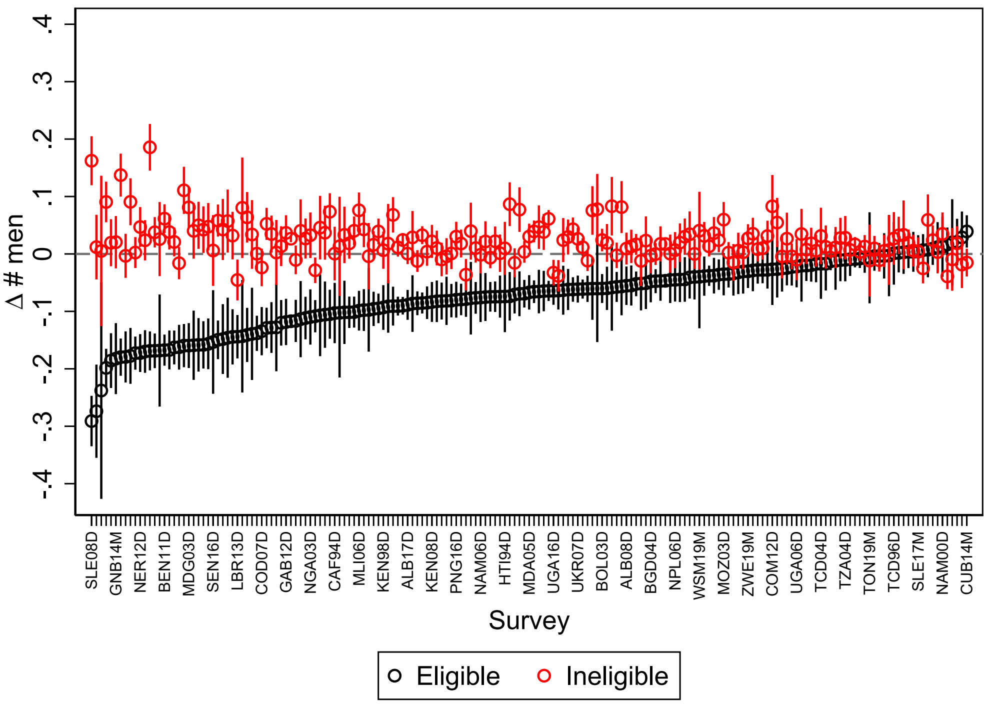


Figure 4: Effect of man's questionnaire on number of eligible and ineligible men

This figure displays estimates of β from equation (3) where the outcome variable is the number of eligible (black) and ineligible men in the household (red), respectively. The sample consists of all 181 DHS and MICS with a man's questionnaire that is randomly assigned across households. Circles indicate point estimates and bars indicate 95% confidence intervals. Surveys are sorted along the x-axis in ascending order of the point estimate on the number of eligible men. Every 5th survey is labelled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively. All estimates are reported in Table A4, columns (2) and (4).

households that are eligible for the man's questionnaire while this is not the case for ineligible households. As a result, household questionnaires in treatment households are more likely to be administered by male data collectors (see Figure A7a). However, consistent with the idea of moral hazard, selection of data collectors into treatment cannot explain the reductions in the number of eligible men as point estimates are barely affected by the inclusion of data collector fixed effects (see Figure A8).[29]

[29]See Appendix A.2 for more details on the effect of the man's questionnaire on data collector characteristics. In the same section, we also examine effects on respondent characteristics.

4.2 Missing women

4.2.1 Empirical strategy

In the DHS and the MICS, women's responses to the woman's questionnaire are of central interest because they are informative about the main focus area of the two survey programs, namely the situation of women and children. Eligible women face substantially longer individual questionnaires than eligible men. In our sample of surveys, the median duration of the woman's questionnaire exceeds the median duration of the man's questionnaire in every single survey. On average, the woman's questionnaire is 16 minutes (64%) longer than the man's questionnaire. In conjunction with the results presented in the previous section, this raises serious concerns about endogenous selection of eligible women.

To assess the amount of missing women of eligible age, we cannot rely on the same identification strategy as for men because in both the DHS and the MICS, the woman's questionnaire is always administered in all households, not just a random subset of households. We identify three (partial) exceptions to this rule, however. In the Ghanaian 2008 DHS, the woman's questionnaire was only administered in a random subset of households. Additionally, in the 2013 DHS in Namibia and the 2019 DHS in Gabon, a short version of the woman's questionnaire was administered to women aged 50 to 64 in a random subset of households (in addition a standard woman's questionnaire for women aged 15-49 in all households). We leverage the random assignment in these three surveys to test if our results for men also hold among women.

We complement this approach with a comparison of the number of female household members of eligible and ineligible age in DHS/MICS and contemporaneous population censuses. This is motivated by the fact that in the DHS and the MICS the number of questions to be administered to women of eligible age (typically aged between 15 and 49) is much larger than the number of questions to be administered to women outside this age range, but no such difference in question load between women of eligible and ineligible age exists in population censuses. This means that data collectors have a strong incentive to omit women of eligible age or to manipulate their age such that they appear to be ineligible in the DHS and the MICS, but they have no such incentive in censuses. Hence, we can compare the average number of women of eligible and ineligible age in the household in the DHS/MICS and the census to test if survey samples contain fewer women of eligible age and (weakly) more of ineligible age.

4.2.2 Data

We form survey-census pairs by matching all DHS and MICS with population censuses conducted within two years of the survey. Since the MICS only records de jure household members, we ensure that censuses matched with MICS record all de jure members.[30] When comparing a DHS to a contemporaneous population census, we restrict the data to de facto household members because DHS statistics generated from individual questionnaires are based on de facto members only.[31] For 77 of the resulting census-survey pairs, we obtain microdata from IPUMS-International (Ruggles et al., 2024) or directly from national statistical offices.[32] See Table A3 for a complete list of the pairs and data sources. They cover 39 countries across Africa, Asia and Latin America, as shown in Figure A10.[33]

To ensure comparability between census and survey data, we exclude collective dwellings from census data. We confirm that the relative question load of eligible to ineligible women is close to one in all censuses, but much larger in the matched DHS and MICS. As shown in Figure A12, the relative question load varies between 1.0 and 1.5 across the matched censuses while it varies between 1.1 to 29.3 across the matched surveys.

4.2.3 Results

Exploiting the random assignment of the woman's questionnaire to households in three DHS, we find a sizable effect of the woman's questionnaire on the presence of eligible women in households in 2 out of 3 surveys - in line with our results for men presented in the previous section. Moreover, the effects of the woman's and the man's questionnaire are of the same order of magnitude within the same survey, as shown in Figure A11.

[30]For a subset of MICS-census comparisons, we restrict the data to de jure members that slept in the household last night because matching censuses do not include de jure members who are absent.

[31]For a subset of DHS-census comparisons, we restrict to de facto members that are usual members of the household because matching censuses do not include de facto members who are visitors.

[32]The authors wish to acknowledge all the statistical offices that provided the underlying data making this research possible. See Table A3 for a complete list of these.

[33]We exclude seven DHS-census pairs where eligibility for the DHS woman's questionnaire is conditional on having ever been married.

Comparing the number of eligible and ineligible women in the household in the DHS/MICS and contemporaneous population censuses paints a similar picture. Figure 5 illustrates that households in the DHS/MICS almost always contain fewer women of eligible age and more of ineligible age. In some cases, however, they contain more or less of both eligible and ineligible women, which points to level differences in the number of recorded household members that may arise from differences in the implementation of household rosters between the DHS/MICS and the census. Importantly, however, the difference in ineligible women between census and DHS/MICS is always at least weakly greater than the difference in eligible women. Thus, in relative terms, the DHS/MICS are under-recording eligible women throughout.

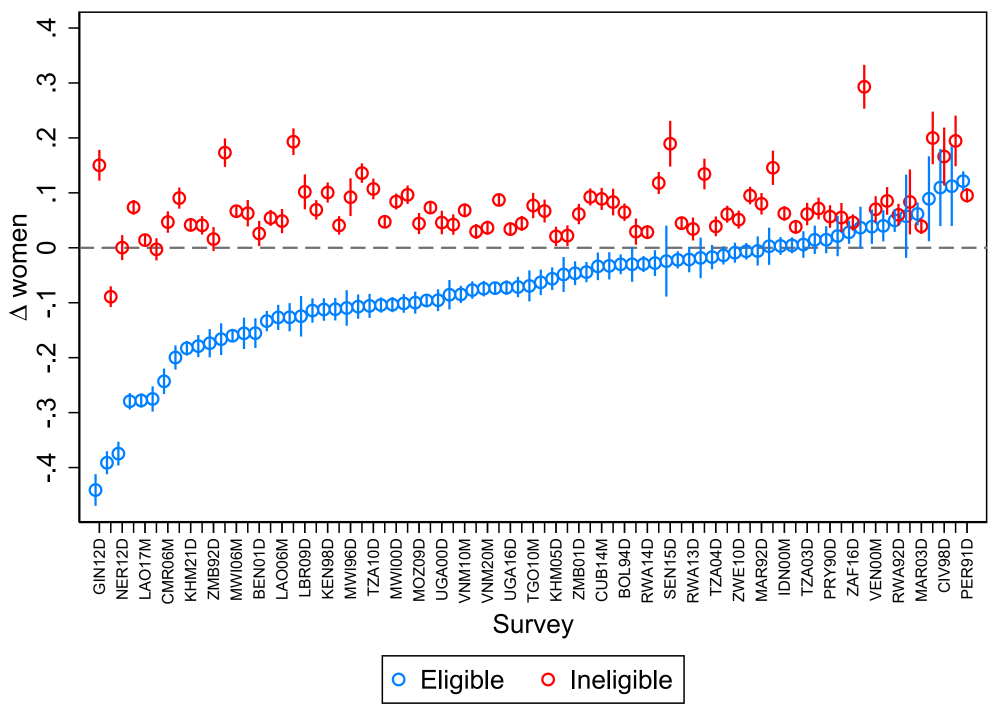


Figure 5: Missing and excess women in DHS/MICS relative to census

This figure displays estimates of β_3 from equation (6) where the outcome variable in the number of women of eligible (blue) and ineligible age (red). The sample consists of all 77 DHS- and MICS-census pairs. Circles indicate point estimates and bars indicate 95% confidence intervals. Surveys are sorted along the x-axis in ascending order of the point estimate on the number of eligible women. Every 2nd survey is labelled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively. All estimates are reported in Table A6, columns (2) and (3).

We provide bounds on the number of missing women that account for potential levels differences in the recording of household members between censuses and

DHS/MICS. We consider two extreme cases. First, to derive a lower bound, assume data collectors do not omit any eligible women from household rosters in the survey. They only engage in manipulation of age to displace eligible women across the eligibility thresholds. In this case, the number of missing eligible women in the survey relative to the census should equal the number of excess ineligible women, and any deviations from this equality would reflect level differences in the recording of household members between survey and census. Hence, the number of missing women is given by half of the difference-in-differences between the number of eligible and ineligible women in survey and census.

Second, to derive an upper bound, assume data collectors do not engage in age displacement. Their only strategy to reduce the number of eligible women they have to interview is to omit such women from household rosters. In this case, any deviation of the difference in the number of ineligible women between survey and census from zero reflects level differences in the recording of household members between survey and census. Thus, the number of missing eligible women is given by the entire difference-in-differences between the eligible and ineligible women in survey and census.[34]

Figure 6 displays the resulting bounds for missing women. We estimate a statistically significantly negative lower bound in 69 out of 77 surveys, ranging between 2% and 17%. In 11 of surveys the lower bound exceeds 10%. The estimated upper bound is substantially larger (in absolute terms) and surpasses 10% in 46 of the surveys. This suggests that a substantial number of eligible women is screened out by DHS/MICS data collectors and never administered the woman's questionnaire. As in the case of men, this appears to be a more serious issue in surveys with longer individual questionnaires (see Figure A13).

To assess the bounds we construct for women, we turn to a subsample of DHS/MICS for which we have both a randomised man's questionnaire and a matched population census. This allows us to compare bounds of missing men for households with a man's questionnaire based on a survey-census comparison with our experimental estimates of the effect of the man's questionnaire. We find that the two approaches yield remarkably similar results (see Figure A14). In 24 out of 33 surveys, the confidence interval of the experimental estimate overlaps with the range of estimates delimited by the bounds. In the remaining cases, the experimental estimate falls short of the lower bound.

[34] See appendix A.3 for details on the estimation of the bounds.

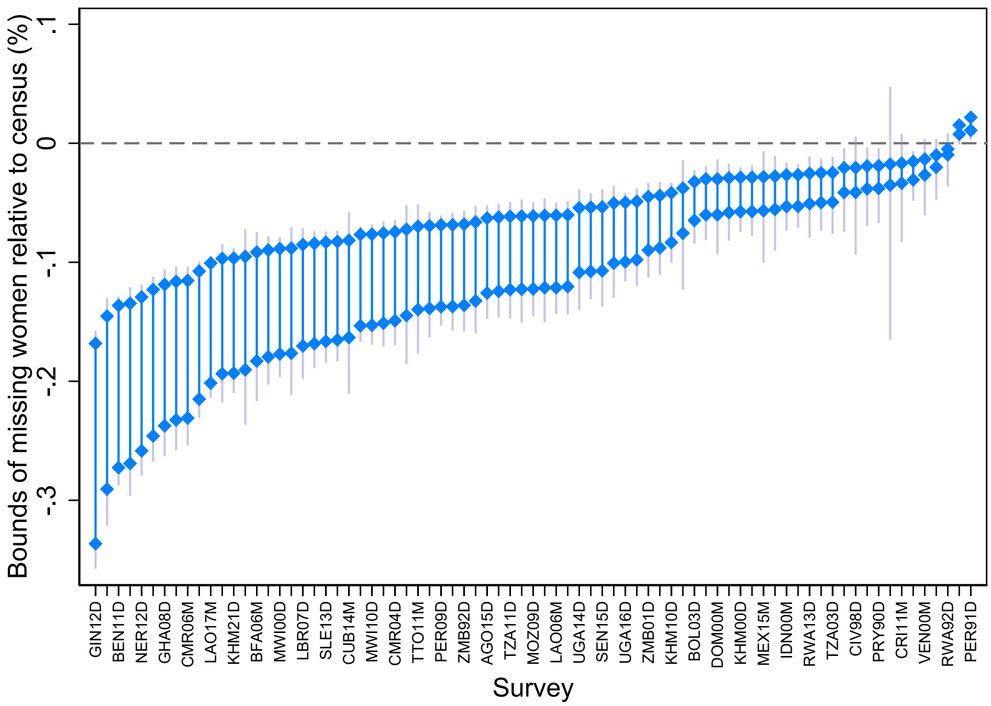


Figure 6: Bounds of missing women in DHS/MICS relative to census

This figure displays estimates of the upper and lower bounds of missing women, indicated by blue diamonds. Grey shaded bars indicate 95% confidence intervals. The sample consists of all 77 DHS- and MICS-census pairs. Surveys are sorted along the x-axis in ascending order of the point estimate of the lower bound. Every 2nd survey is labelled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively. All estimates are reported in Table A6, columns (4) and (5).

To facilitate the comparison of our estimates across surveys and sex, we normalize our estimates of missing men and women relative to the length of the individual questionnaires faced by eligible household members. To this end, we compute the elasticity of the number of recorded eligible household members with respect to question load. We define this elasticity as the relative reduction in the number of eligible household members over the relative increase in question load for eligible household members. We measure the question load by the total number of questions listed in the questionnaires that a household member is eligible for (household roster, man's questionnaire, woman's questionnaire).[35]

[35]More formally, the elasticity is $\varepsilon = (\beta/m_c)/[(q_{eligible}/q_{ineligible}) - 1]$ where β is the effect of household assignment to the man's questionnaire on the number of eligible men as estimated in equation (3), m_c is mean number of eligible men in control households and q_i is the number of questions that an (in)eligible member has to be asked. See Appendix A.1.3 for details on the counting of questions listed in questionnaires.

We find similar elasticities for men and women. The elasticity for men estimated from the random assignment of the man's questionnaire is on average −0.010, with variation across surveys between −0.001 at the 10th percentile and −0.021 at the 90th percentile. The elasticity for women estimated from the survey-census comparison ranges between −0.002 at the 10th percentile and −0.027 at the 90th percentile, with an average of −0.008. In surveys where we can estimate both of these elasticities (33), they are quantitatively similar and rarely statistically significantly different from each other, as shown in Figure A15.[36]

4.3 Mechanisms

4.3.1 Effort cost

The theoretical framework laid out in section 2 suggests that endogenous sample selection is a bigger concerns if the effort cost of recording household members is larger. We test this prediction using variation in temperature and humidity as a source of exogenous variation in effort cost. Motivated by experimental evidence of their negative impact on human physiology and performance (Pilcher et al., 2002; Seppanen et al., 2005; Cui et al., 2013), we hypothesize that reductions in eligible men will be more pronounced at low and high temperatures relative to intermediate temperatures.

We use wet bulb temperature as our preferred measure of temperature, following the recent literature in economics (Geruso & Spears, 2018; Adhvaryuy et al., 2019; LoPalo, 2023). Wet bulb temperature accounts for relative humidity which interacts with temperature in the generation of heat stress. Note that it is lower than dry bulb temperature unless relative humidity is 100% – in which case the two temperature measures are equal. Moderate conditions, such as a dry bulb temperature of 75°F and a humidity of 40%, correspond to wet bulb temperatures of around 60°F. Wet bulb temperatures of 80°F capture extreme heat, corresponding to dry bulb temperatures of 100°F at 40% humidity or 90°F at 65% humidity.

We estimate how the effect of household assignment to the man's questionnaire varies with temperature changes within survey cluster and data collector, adopting a semi-parametric specification to allow for non-linearities in the effect of temperature:

[36] See Figures A16 and A17 for the full distribution of elasticity estimates for men and women.

$$y_{icdrt} = \sum_j \beta_j (T^j_{ct} \times MQ_{ic}) + \sum_j \gamma_j T^j_{ct} + \eta Precip_{ct} + \mu_c + \theta_d + \lambda_r + \epsilon_{icdrt} \quad (4)$$

where y_{icdrt} is the number of eligible men in household i interviewed on the r-th day of data collection in survey cluster c by data collector d on date t. MQ_{ic} is an indicator variable that takes value one if household i in cluster c is eligible for the man's questionnaire, and zero otherwise. T^j_{ct} is an indicator that takes the value one if the daily average wet bulb temperature in cluster c on date t falls into temperature bin j; where we consider the following temperature bins: $< 40°F$, $40°F - 50°F$, $50°F - 60°F$, $60°F - 70°F$, $70°F - 80°F$, $> 80°F$. The coefficients of interest are β_j capturing the reduction in the number of eligible men due to household assignment to the man's questionnaire in the different temperature bins. We control for precipitation $Precip_{ct}$, survey cluster fixed effects μ_c, data collector fixed effects θ_d and fixed effects for the day of data collection in the survey cluster λ_r. Standard errors are clustered at the survey cluster level.

We construct the daily average wet bulb temperature in each survey cluster on each survey date following LoPalo (2023). We source global daily weather information for each 0.25 degree latitude/longitude increment from the Princeton Meteorological Forcing Dataset.[37] For each survey cluster, we set the wet bulb temperature to the average across the four surrounding grid points, weighting by the inverse distance between the cluster and each grid point. This way, we link 17,716 survey clusters from 49 of the surveys in our sample to weather information. Clusters from the remaining surveys cannot be matched for two reasons. First, the weather data data only covers the time period until 2010. Hence, surveys post 2010 cannot be matched. Second, GPS coordinates of survey clusters are not available.

In line with our hypothesis, we find an inverse U-shaped relationship between the treatment effect and wet bulb temperature, as shown in Figure 7. Treatment effects are smallest at wet bulb temperatures between 50 and 70 where 0.08-0.09 eligible men are estimated to be missing. We find substantially more missing men at higher temperatures. Approximately 0.12 eligible men are missing at temperatures between 70 and 80, and 0.15 at temperatures above 80. We also observe somewhat more missing men at temperatures below $50°F$, although estimates are more noisy.

[37]The original data is 3-hourly. We work with the average wet bulb temperature across the 8 daily readings. For more details on the weather dataset see LoPalo (2023).

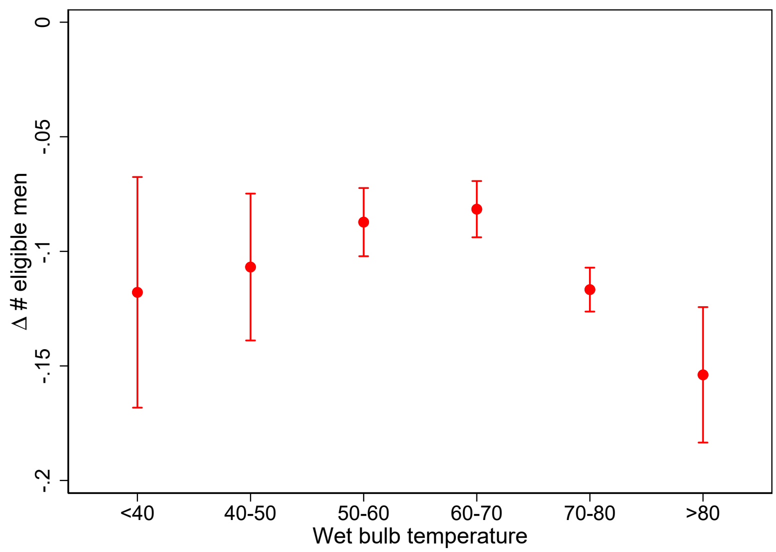


Figure 7: Missing men by wet bulb temperature bin

This figure displays the estimates of the regression coefficients β_j from specification (4) across temperature bins. Bars indicate 95% confidence intervals. Standard errors are clustered at the survey cluster level.

Interestingly, the U-shape we estimate is similar to the relationship between ambient temperature and human productivity estimated by others (Cai et al., 2018; LoPalo, 2023), suggesting that data collectors partially compensate for productivity losses experienced at high temperatures by reducing the number of eligible household members. More broadly, our findings indicate that higher temperatures may aggravate moral hazard problems in the workplace.

4.3.2 Detection probability

Another prediction from the theoretical framework is that screening out of eligible men will be less severe if the probability of getting caught doing so is higher. To examine this prediction, we study the correlation between the implementation of systematic backchecks, also referred to as mandatory re-interviewing, and the share of missing eligible men across surveys.[38] Consistent with the theory, we find that the share of missing men is lower in surveys that conduct systematic backchecks (see Figure A18). In surveys with backchecks, the average share of missing eligible men amounts to 4.5% compared to 7.3% in surveys without backchecks.

[38]We manually code whether or not each survey featured mandatory re-interviewing based on the final survey report.

5 Selection

Eligible men and women imply a high effort cost for data collectors because they have to be administered lengthy individual questionnaires. In the previous section, we have established that in many DHS and MICS a significant number of these men and women are excluded from individual questionnaires as a result of manipulation by data collectors. Consequently, these men and women are missing from the database underlying all statistics based on information collected in individual questionnaires. These statistics, however, are the core output of the DHS and the MICS program. They include crucial information on topics such as fertility, maternal health, HIV, marriage and domestic violence. This raises two questions. First, do the excluded men and women differ systematically from the included ones? Second, if so, to what extent does this lead to bias in important aggregate statistics?

5.1 Who is missing?

Who are the eligible household members excluded from individual questionnaires by data collectors? 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 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. We examine these differences running specification (3) on individual-level characteristics recorded in the household roster (and thus observable for all men, independent of their household's eligibility for the man's questionnaire).

For women, we simply compare the average characteristics of eligible women in the DHS/MICS to those in the census. To this end, we harmonise information on age, the relationship to the household head, years of schooling and marital status between DHS/MICS and censuses as detailed in Section A.1.4.

We find that missing men and missing women differ systematically from included ones in remarkably similar ways. First, in most surveys, eligible household members in treatment households are older than eligible household members in the respective comparison group (Figure A19), indicating that younger members are more likely to be screened out. We show that this results from the combination of two facts, focusing on the case of men. First, eligible men that are within 10 years of age from the lower and upper eligibility thresholds (in most surveys 15-24 and 40-49 years old) are about equally likely to be screened out of the sample for the man's questionnaire, and twice as likely as eligible men who are further in age from these thresholds (typically 25-39 years old), as shown in Figure A20. Second, in most surveys, the age distribution peaks at young ages, which implies that more young people are screened out.[39]

We further find that in most surveys, eligible household members in treatment households are more closely related to the head of their household, more educated and more likely to have ever been married relative to the corresponding comparison group (see Figure 8).[40] This implies that missing men and women tend to be less closely related to the head of their household, less educated and less likely to have ever been married. One interpretation of these findings is that data collectors predominantly screen out eligible individuals that are at the periphery of their respective households. These are precisely the household members where data collectors have discretion because household definitions are sufficiently vague, with rosters typically instructing data collectors to list all 'usual members' (plus visitors that slept in the household last night in the case of the DHS). Moreover, omission or age manipulation are plausibly less likely to cause opposition from respondents or supervisors in these cases - all of whom also have an incentive to keep surveys short. Stated in the terms of our theoretical framework in section 2, the probability of detection is likely to be lower.

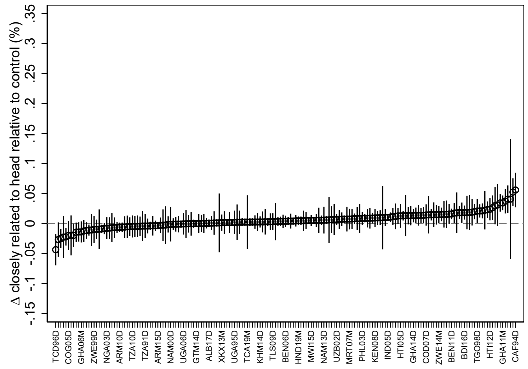
Reassuringly, we find few signs of selection in surveys for which we estimate little missingness. As Figures A21 and A22 show, in surveys with few missing eligible (wo)men, eligible wo(men) in treatment (survey) and control (census) households look similar on observables. In surveys with more missing eligible household members, differences tend to be larger.

Complementary evidence suggests that the missing household members often belong to marginalised populations. In a subset of the DHS, we observe additional individual characteristics in the household roster for specific age groups. Pooling the data across all surveys with available data, we find that eligible men in households assigned to the man's questionnaire are 10% less likely to be poor (ages 15-17), 5% less likely to be chronically sick (ages 18-59), and equally likely to disabled (all ages) or orphaned (ages 15-17) – see Figure A23.[41]

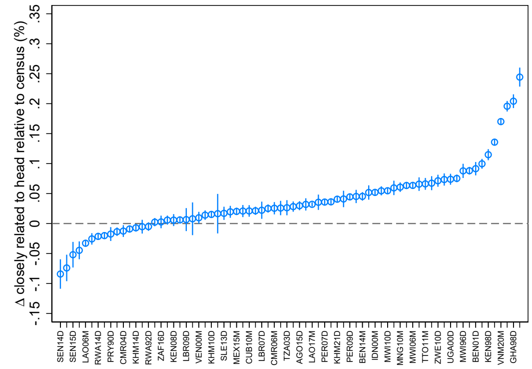
[39] It is also remarkable that even in the intermediate age range, far from the eligibility thresholds, more than 5% of men are missing in some surveys (14).

[40] All estimates are reported in Tables A8 and A9.

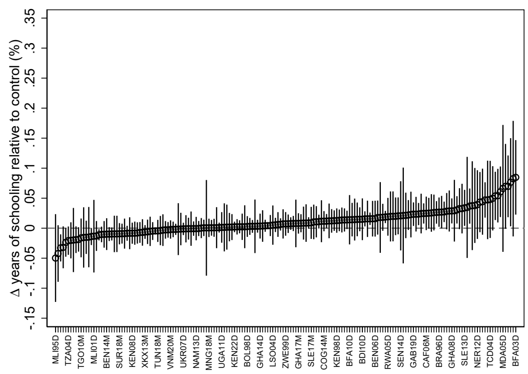
[41] See appendix A.1.4 for details on the underlying samples and variable definitions.



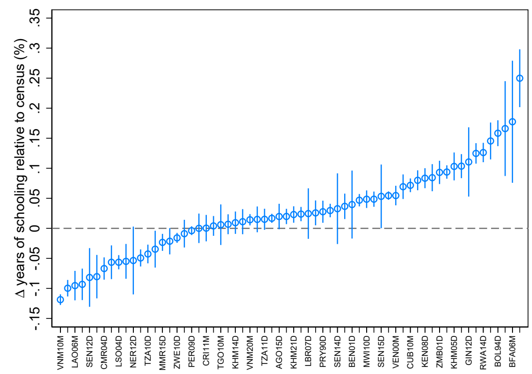
(a) Relationship to head: Men



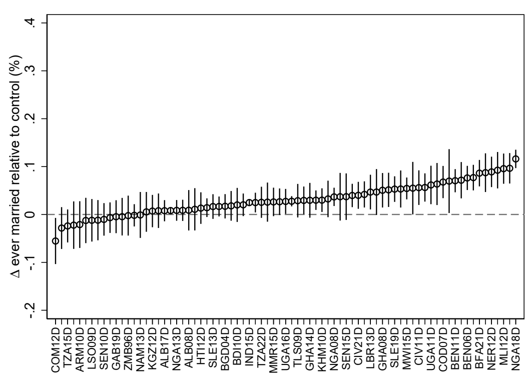
(b) Relationship to head: Women



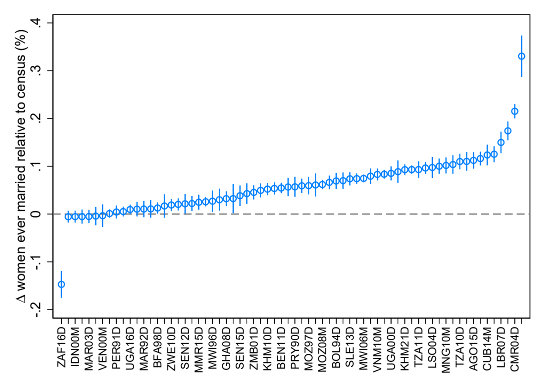
(c) Years of schooling: Men



(d) Years of schooling: Women



(e) Marital status: Men



(f) Marital status: Women

Figure 8: Selection on observables

This figure displays estimates of the effect of household assignment to the man's (left) and woman's questionnaire (right) on the characteristics of eligible men and women relative to the relevant comparison group, i.e., control households for men and census households for women. See Section A.1.4 for details on the construction of all outcome variables. Standard errors are clustered at the household-level. Circles indicate point estimates and bars indicate 95% confidence intervals. Surveys are sorted along the x-axis in ascending order of the respective point estimate. In panels (a) and (c) every fifth survey is labeled, in all other panels every 2nd survey is labeled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively. All estimates are reported in Tables A8 and A9, columns (3)-(5).

5.2 Bias in aggregate statistics

How does the selective screening out of household members revealed in the previous section affect aggregate statistics? The documented selection on observables implies that endogenous sample selection does not only lead to a decline in precision of estimates as a result of sample size reductions. It also leads to bias in aggregate statistics. How important is this bias? In this section, we address this question focusing one key survey outcome – fertility.

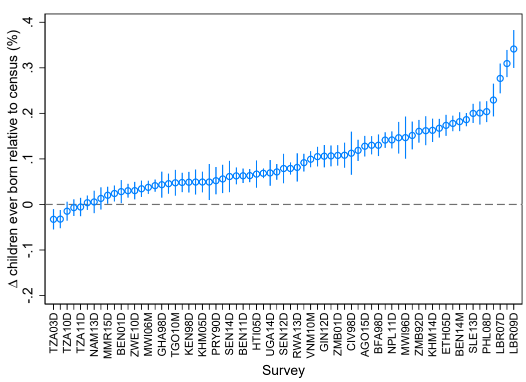
The DHS and the MICS are a key data source on fertility in low- and middle-income countries. The large number of top demography papers citing the DHS and the MICS is evidence of this. Between 2013 and 2017, 15.4% of all papers published in the two top journals *Demography* and the *Population and Development Review* cited the DHS or the MICS.[42] Work on fertility in the field of Economics also heavily relies on the two household survey programs (Vogl, 2016; Chatterjee & Vogl, 2018; Rossi, 2018; Dupas et al., 2023; Zipfel, 2024). Additionally, fertility data from the two programs is a key input for national health, family planning and education programs, not least due to the weakness of vital registration systems in large parts of the world. In fact, the DHS and the MICS are considered the only reliable data source on fertility in many contexts.

We focus on the total number of live births as our measure of fertility because this information is most consistently gathered in the two survey programs as well as population censuses. We observe the total number of children ever born to eligible women in 67 out of 77 survey-census pairs, i.e., in both the survey (from the woman's questionnaire) and the matched population census.[43] Comparing the number of reported live births within pairs, we find evidence of significantly higher fertility in DHS/MICS than in contemporaneous censuses. Figure 9a shows that the average number of children ever born in the surveys exceeds the one in the census in 56 out of 61 cases. In 38 of these cases, the gap is larger than 5%, and in 26 of them larger than 10%. Only in two cases, we detect a statistically significantly lower reported fertility in DHS/MICS than in the census.[44] Reassuringly, these are surveys where we only find limited evidence of missing women. Overall, the degree of overestimation is strongly negatively correlated with our estimates of missing women (see Figure A24).

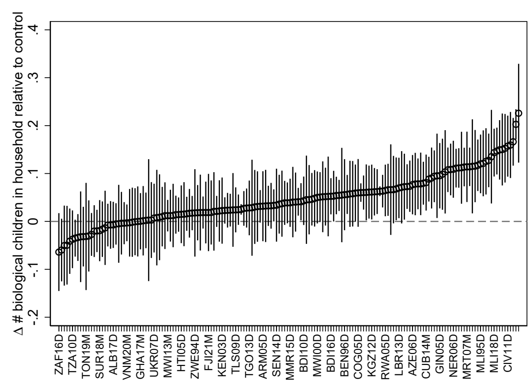
[42] See Appendix A.4 for details.
[43] See Appendix A.1.4 for details on the survey-census harmonisation of this information.
[44] All estimates are reported in Table A9, column 6.

Complementary evidence from the random assignment of the man's questionnaire corroborates the overestimation of fertility in the DHS and the MICS due to endogenous sample selection. In the absence of information on the fertility of men in households without a man's questionnaire, we show that the number of biological children men live with in their household, is larger in treatment households in the majority of surveys.[45] As Figure 9b shows, the point estimate is positive for 93 out of 117 surveys, and statistically significantly so in 42. On average, fertility is overestimated by 4% and in 24 surveys, overestimation exceeds 10%.

In conclusion, our findings suggest that data collectors are more likely to screen out eligible men and women with less children.[46] Ultimately, this leads to a significant upward bias in fertility measures in the DHS and the MICS.



(a) Fertility of women (census comparison)



(b) Fertility of men (DHS/MICS RCT)

Figure 9: Bias in aggregate fertility statistics

This figure displays estimates of the effect of household assignment to the man's (left) and woman's questionnaire (right) on the measures of fertility of eligible men and women relative to the relevant comparison group, i.e., control households for men and census households for women. See Section A.1.4 for details on the construction of all outcome variables. Standard errors are clustered at the household-level. Circles indicate point estimates and bars indicate 95% confidence intervals. Surveys are sorted along the x-axis in ascending order of the respective point estimate. On the left, every 2dn survey is labeled, on the right every 5th survey is labeled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively. All estimates are reported in Tables A8 and A9, column (6).

[45] Since the fertility of men is only elicited in the man's questionnaire, we do not observe fertility of men in control households. We overcome this limitation by constructing a proxy of fertility of men in both treatment and control households from the parent survival module in the household roster. This module is included in 168 out of the 181 surveys in our sample and links children aged 17 and younger to their biological parents as long as these are alive and live in the same household. Thus, we can compute the number of biological children each eligible man lives with. To obtain nationally representative figures, we weight households using their sampling weights.

[46] Note that this is in line with our result in Subsection 5.1 showing the relative absence of never married eligible men and women from households that are assigned to individual questionnaires.

6 Applications and implications

We show above that endogenous sample selection leads to significant numbers of missing subjects in human-collected data. Resulting samples are selected and marginalised populations are more likely to be missing. Is endogenous sample selection of broader concern beyond affecting aggregate statistics? In this section, we provide three applications showing how endogenous sample selection in human-collected data can systematically alter statistical inference and bias economic analysis.

First, endogenous sample selection can be correlated with a given treatment, in particular if the treatment affects the effort cost of data collectors. In such a case, even otherwise well-identified experimental or quasi-experimental estimates will be biased. In Subsection 6.1, we demonstrate that climate shocks in Sub-Saharan Africa are correlated with sample selection in the DHS, thereby leading to spurious relationships between shocks and outcomes of interest.

Second, to study labour and capital market frictions researchers often use firm or farm size distributions for inference, either explicitly by testing for bunching at policy-relevant thresholds or implicitly by leveraging moments of the firm size distribution to estimate model parameters. However, firm and farm censuses frequently employ size thresholds to determine sample inclusion and question load. In Subsection 6.2, we show that endogenous sample selection leads to distortions in recorded firm-size distributions. In particular, the entire 'missing middle' of firms in the Indian Economic Census can be explained by data collector incentives, with direct implications for structural estimation that uses these distorted firm-size distributions.

Third, endogenous sample selection can introduce initial selection of subjects into longitudinal surveys, generating dynamics in outcomes that are not representative of underlying population dynamics. In Subsection 6.3, we show that high-effort-cost individuals are missing from the US National Longitudinal Survey of Youths 1997 (NLSY97) and that the missing appear positively selected on family income. As a result, the dynamics of youth employment are diverging from those estimated from comparable survey data, for example with respect to gaps across gender.

6.1 The impact of climate shocks in Africa

A rapidly growing literature examines the impact of climate shocks on a wide range of outcomes, including health (Burke et al., 2015; Fichera & Savage, 2015; Nagata et al., 2021; Le & Nguyen, 2022), mortality (Geruso & Spears, 2018), fertility (Norling, 2022), marriage (Corno et al., 2020; Corno & Voena, 2023), domestic violence (Epstein et al., 2020), consumption (Paxson, 1992; Dimitrova, 2021), wealth (Thiede, 2014) and conflict (Miguel et al., 2004; Couttenier & Soubeyran, 2014).[47] In this literature, climate shocks are typically considered exogenous events and associations between shocks and outcomes are interpreted causally. However, if outcome data is collected by humans on the ground, selection into the sample can itself be affected by climate shocks insofar as these shocks impact the behaviour of data collectors.[48] This, in turn, undermines the causal interpretation of the association between shock and outcome even if the shock is truly exogenous.

To assess the relevance of this concern, we examine how rainfall shocks in Sub-Saharan Africa affect the extent to which data collectors screen out eligible men in eligible households. To this end, we assign all geo-referenced DHS clusters in our data to 0.5×0.5 arc degree grid cells.[49] Using gridded rainfall data from the Climate Hazards group Infrared Precipitation with Stations (CHIRPS) version 2.0 dataset, we construct annual rainfall for each grid cell-year in our data. Following Burke et al. (2015) and Corno et al. (2020), we define a drought as calendar year rainfall below the 15th percentile of a grid cell's long-run rainfall distribution. Analogously, we define a flood as calendar year rainfall above the 85th percentile of a grid cell's long-run rainfall distribution. Using the entire CHIRPS time series from 1981 until 2024, we fit a gamma distribution of calendar year rainfall for each grid cell. Then, we use the estimated gamma distribution for a given cell to assign each calendar year's rainfall realisation to its corresponding percentile in the distribution.

We estimate how household assignment to the man's questionnaire interacts with changes in weather conditions within grid-cell using the following regression model:

$$y_{igct} = \beta_0 + \beta_1 MQ_{igct} + \beta_2 Shock_{gt} + \beta_3 (MQ_{igct} \times Shock_{gt}) + \mu_{cg} + \tau_{ct} + \epsilon_{igct} \quad (5)$$

where y_{igct} is the number of eligible men in household i in grid cell g in country c in year t, MQ indicates household assignment to the man's questionnaire and $Shock$ is

[47] See Carleton and Hsiang (2016) for a summary of the growing literature on the social and economic impacts of climate.

[48] Climate shocks can affect data collector behaviour in various ways, for example by changing effort cost of collecting data as shown in Section 4.3, thus affecting effort levels; or by affecting labour market conditions that alter collectors' outside options and, thus, effort levels on the job.

[49] At the equator, this corresponds to an area of approximately 2,500 square kilometers.

an indicator for drought or flood.

We find that climate shocks affect sample selection. As shown in Table 1, eligibility for the man's questionnaire reduces the number of eligible men by approximately 1/3 less during droughts and increases it by about 1/5 more during floods. This has repercussions for estimated relationships between rainfall shocks and outcomes derived from the DHS, such as marriage and fertility. In the sample of households that are eligible for the man's questionnaire, we observe a significantly negative association between marriage of eligible men and dryness – presence of droughts and absence of floods – while we do not detect a statistically significant association among control households. Similarly, the negative relationship between dryness and fertility, as proxied by the number of biological children of eligible men in the household below the age of one, is much stronger in the sample of treatment households. In other words, the composition of the sample of eligible men in treatment households is more sensitive to rainfall shocks than the one in control households, leading to spurious relationships between shocks and outcomes.[50]

In conclusion, the reported effects of climate shocks on sample selection caution against a causal interpretation of correlations between climate shocks and contemporaneous outcome data collected by humans in the field.

6.2 The firm size distribution in India

Firm and farm size distributions are commonly used to study labour and capital market frictions and factor misallocation. However, many of the underlying firm and farm censuses use size thresholds to determine sample inclusion or question load. In particular, many censuses limit the amount of information collected about small units, thus incentivizing data collectors to manipulate unit size such that units fall below the size threshold or to omit larger units entirely.

The Indian Economic Census is an important example of such a data collection design. It is heavily used in economics, featuring in at least six top general interest publications over the last five years alone. Yet, its design creates an incentive for data collectors to manipulate firm size. It aims to record all formal and informal non-farm businesses in the country. To this end, data collectors visit all buildings in the entire

[50] Note that the sample is changing across columns in Table 1 because the measures of marriage and fertility are not available in all surveys. For completeness, Appendix Table A10 shows the effect on the number of eligible men for each sample separately.

Table 1: Interaction of extreme rainfall events with question load

	# eligible men		married		# children below age 1	
	(1)	(2)	(3)	(4)	(5)	(6)
Man's Questionnaire	-0.091***	-0.084***	0.027***	0.024***	0.011***	0.009***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
Drought	0.005		0.002		-0.004*	
	(0.011)		(0.005)		(0.002)	
Drought x MQ	0.030***		-0.015***		-0.005***	
	(0.008)		(0.004)		(0.002)	
Flood		0.005		-0.006		0.001
		(0.010)		(0.005)		(0.002)
Flood x MQ		-0.018**		0.007*		0.004*
		(0.007)		(0.004)		(0.002)
Constant	1.034***	1.034***	0.511***	0.512***	0.103***	0.102***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)
Number of surveys	73	73	47	47	63	63
Observations	865,214	865,214	656,743	656,743	825,861	825,861
R^2	0.109	0.109	0.032	0.032	0.018	0.018

All regressions include country-grid cell fixed effects and country-year fixed effects. Regressions in columns (1) and (2) are at the household level, regressions in the remaining columns at the individual level. In columns (3) through (6), the sample is restricted to men of eligible age. Married is an indicator variable that takes value 1 if a man is married or living with their partner, and zero otherwise. The number of children below age 1 captures the number of biological children of a man that are aged 0 to 11 months and live in the same household as the man. MQ is an indicator variable that takes the value one if a household that is eligible for the man's questionnaire, and zero otherwise. Drought and flood events are defined as described in the text. Standard errors are clustered at the country-grid cell level.
* p<.10, ** p<.05, *** p<.01

country, recording the firms found therein and their basic characteristics, including the total number of employees. Subsequently, additional information is collected for firms above a given size threshold.

We exploit the shifting of the eligibility threshold for additional data collection between consecutive censuses to reveal firm size manipulation by data collectors. In fact, in 1998 no such threshold existed – the amount of information collected about firms was independent of their size. In the 2005 Economic Census, the requirement to complete an address slip for all firms employing ten or more workers was introduced (Ministry of Statistics and Programme Implementation, India, 2005). In the next firm census in 2013, the eligibility threshold was adjusted downward to a firm size of eight and the additional information requirement was expanded to include a description of a firm's major activity and its source of registration alongside their name and address. All of this information was collected on a form labelled 'Schedule 6C'.

Figure 10 illustrates bunching of firms below the respective eligibility thresholds of ten and eight in 2005 and 2013, and does not reveal any sign of bunching in the 1998 firm size distribution. In fact, the right panel for 1998 shows that the firm size distribution in India closely follows a power law, a regularity that has previously been observed in many other countries (Axtell, 2001; Hernández-Pérez et al., 2006). In contrast, the 2005 and 2013 distributions clearly display excess mass to the left of their respective threshold and missing mass to the right of it. Moreover, bunching moves in accordance with the change in threshold between these two years. We interpret this as evidence of firm size manipulation by data collectors.[51] Fitting a linear model to describe the relationship between the log share of establishments and the log number of employees (as shown in right-hand side panels of Figure 10), we estimate that the observed manipulation reduced recorded employment by approximately 6.3 and 2.8 million workers in 2005 and 2013, respectively.[52]

Our findings have implications for three separate streams of literature. First, they suggest a novel explanation for the much debated phenomenon of a 'missing middle' in the firm size distributions of low- and middle-income countries. Minimum thresholds for the collection of detailed firm data create an incentive for data collectors to adjust recorded firm size downwards, generating missing mass in the middle of the distribution. Hence, differences in the design of data collection across countries as well as data sources within countries may help explain seemingly contradictory findings in the literature (Tybout, 2000, 2014; Hsieh & Olken, 2014; Abreha et al., 2022).

Second, our findings have implications for structural work on labour and capital market frictions because this literature frequently uses moments of the firm size distribution for calibration or structural estimation. The firm size distribution in the Indian Economic Census, for example, has been leveraged to study labour regulation (Amirapu & Gechter, 2020), microfinance (Buera et al., 2020) and female entrepreneurship

[51] We are not aware of any other incentive structures for data collectors or firm owners that have changed in the same way. Admittedly, there is a range of regulations in India limited to firms of size ten and above (Amirapu & Gechter, 2020). But while these may contribute to the bunching at ten observed in 2005, they cannot explain the drastic reduction in firms of sizes eight and nine between 2005 and 2013, leading to a relative lack of firms of these sizes in the later year, because the regulatory threshold remained ten.

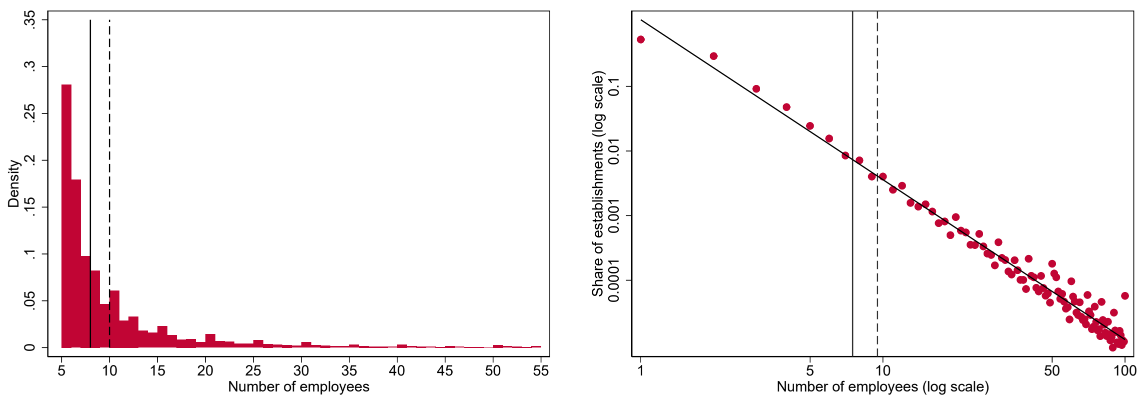
[52] We assume that the manipulation window ranges from seven to 30 in 2005 and from five to 28 in 2013. Our estimates increase if we allow the manipulation window to be larger. We exclude firms with less than two workers and more than 90 workers when fitting the model. We choose the latter threshold because firms with more than 100 workers are subject to differential regulation which could affect the firm size distribution around this firm size.

(Chiplunkar & Goldberg, 2024). Hence, the distortions in the firm size distribution introduced by data collectors as well as the changes in these distortions over time directly affect this line of work.

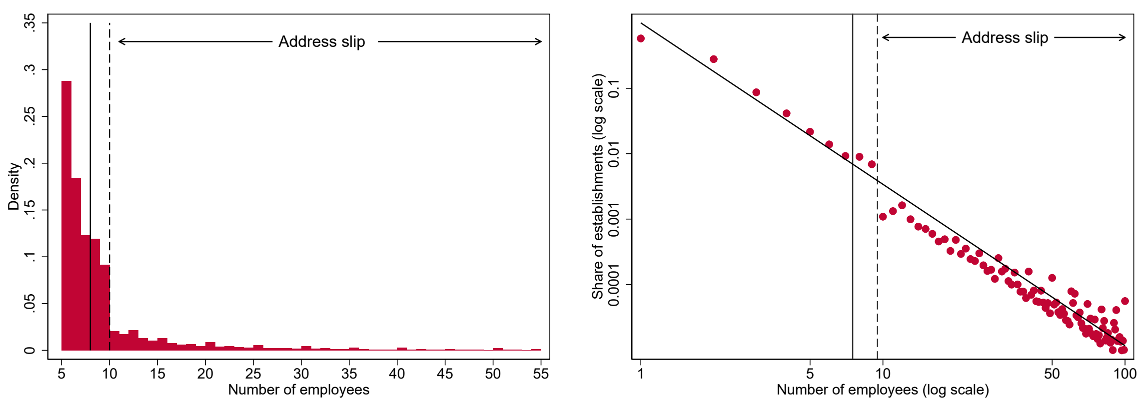
Third, the uncovered manipulation of firm size by data collectors has implications for reduced form work evaluating effects on non-farm employment, as recorded in the Economic Census. This includes recent work on the impact of roads (Asher & Novosad, 2020), electricity (Burlig & Preonas, 2024), canals (Blakeslee et al., 2023; Asher et al., 2024), public employment programs (Muralidharan et al., 2023) and politics (Asher & Novosad, 2017). If the treatment under study is uncorrelated with the share of non-farm workers that are missing due to firm size manipulation, treatment effects will be attenuated because recorded non-farm employment is less sensitive to treatment than actual non-farm employment.[53] If treatment is instead correlated with the share of non-farm workers that are missing due to firm size manipulation, then treatment effects can be upward or downward biased depending on whether treatment makes firms more or less susceptible to downward size manipulations by data collectors. For example, treatment could generate growth among initially small firms, thereby shifting a lot of these above the size threshold. Data collectors, however, would adjust their recorded employment downwards, thus (partially) masking the growth effects of treatment.

In short, the example of the Indian Economic Census illustrates that manipulation of firm size by data collectors in response to incentives embedded in data collection protocols has far-reaching consequences across a wide realm of economic research.

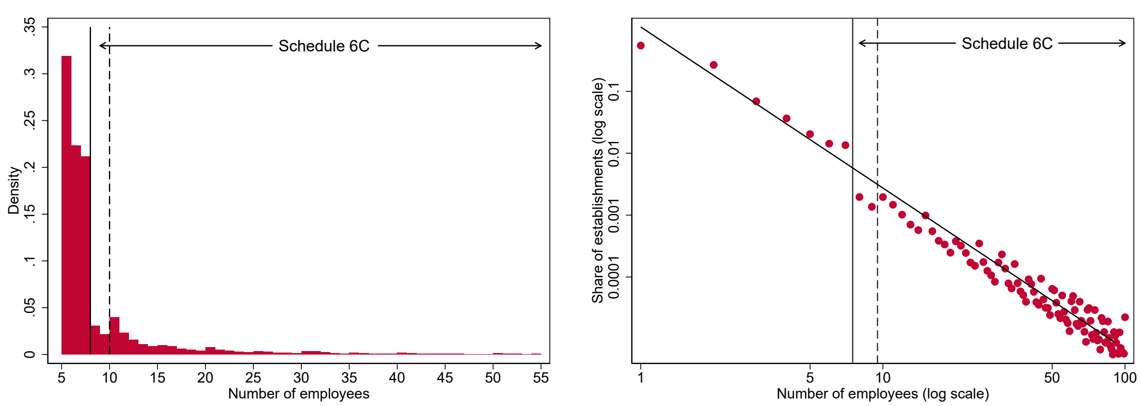
[53] Consider the evaluation of a randomly assigned treatment T on non-farm employment E. Let s be the share of missing workers. The estimated treatment effect β is the difference in mean recorded employment $\bar{E_R} = \bar{E} * (1-s)$ between treatment and control: $\beta = \bar{E_R^T} - \bar{E_R^C} = (1-s)(\bar{E^T} - \bar{E^C})$.



(a) Economic Census 1998: no addtional firm-size dependent schedules



(b) Economic Census 2005: additional schedule for firms with 10+ employees



(c) Economic Census 2013: additional schedule for firms with 8+ employees

Figure 10: Firm size distribution in the Indian Economic Census

This figure plots the firm size distribution in the 1998, 2005 and 2013 Indian Economic Census. The panels on the right show the distributions in levels, the panels on the left in logs. Vertical lines indicate firm size thresholds above which data collectors had to complete additional schedules in 2005 (dashed line) and 2013 (solid line). Note that in 1998, there was no variation in the number of schedules to be completed by firms size.

6.3 The gender gap in youth employment in the US

Endogenous sample selection is a particular concern in the recruitment of participants for longitudinal studies because the initial selection into participation affects the representativeness of all subsequent survey rounds. Moreover, standard static re-weighting on baseline observables may not be sufficient to recover representative dynamics over time.

The National Longitudinal Survey of Youth 1997, also referred to as the NLSY97, is a case in point. The NLSY97 is a panel survey that has re-interviewed a sample of American youth 20 times since 1997. It is heavily used in economics, in particular in the field of labour, featuring in at least thirteen top general interest economics papers published since 2010. The process to identify and recruit study participants, however, created an incentive for data collectors to screen out eligible youth.

The initial sample of the NLSY97 was designed to be representative of the civilian non-institutionalised population of cohorts born between 1980 and 1984. First, using standard area-probability sampling methods, primary sampling units (PSUs) were randomly drawn. Second, addresses were randomly selected within each PSU. Third, face-to-face screening interviews were conducted at all selected addresses to identify eligible youths in these households and recruit them for the study.

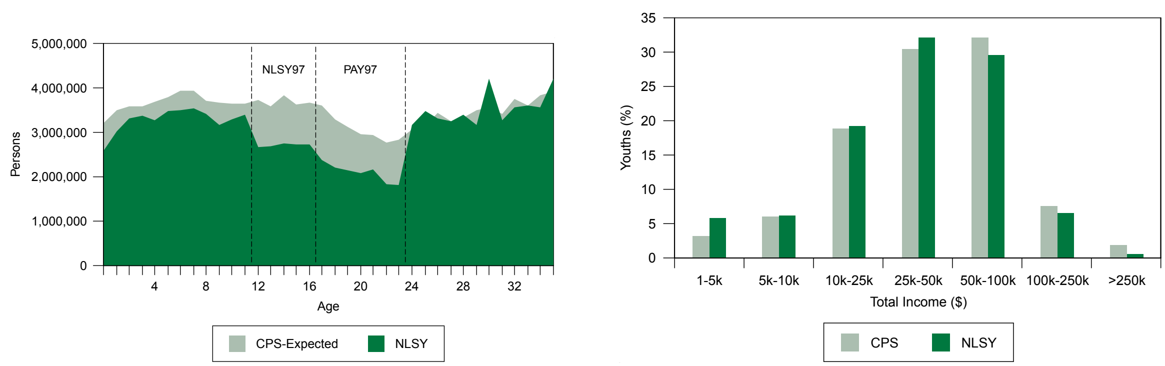
Screening interviews for the NLSY97 were conducted in 1997 and 1998. The same interviews were also leveraged to screen for eligible participants for the Profile of American Youth (PAY97), the second round collection for research on vocational aptitude of US youth. All youth aged 12 to 16 as of December 31st, 1996 were eligible to participate in the NLSY97, while those aged 18 to 23 were eligible for the PAY97.

The screening procedure was as follows. First, data collectors conducted a brief screener with a household informant aged 18 or above. If any youth eligible for the NLSY or PAY were identified, data collectors continued by gathering detailed information about all household members. This extended screener was used to confirm eligibility for the two studies. In households with eligible youth, data collectors then moved on to the first round interviews, including several detailed parent and youth questionnaires.[54] This 'screen-and-go' procedure created an incentive for data collectors *not* to identify eligible youth because households with eligible youth were a lot more work for them.

[54]If the corresponding respondents were not available during the visit, a follow-up was scheduled.

Examinations of the screening results reveal a pronounced undercoverage of youth of eligible ages (Horrigan et al., 1999; Moore et al., 2000). As Figure 11a shows, screened households display a lot fewer youth aged 12 to 23 than comparable households in the March 1997 CPS. Undercoverage at younger age is much more pronounced than at older ages – in fact at age 24, there is overcoverage in NLFS97, if anything. This pattern is strikingly similar to the one observed in Figure 1 and entirely consistent with endogenous sample selection.[55]

Moore et al. (2000) also provide evidence of selection on observables. In particular, they show that the family income of youth in the NLSY97 is lower than the family income of households with resident youth aged 12 to 16 in the CPS – see Figure 11b.[56]



(a) Age distribution in screened households (b) Distribution of youth by family income

Figure 11: Sample selection in the National Longitudinal Survey of Youth 1997

The left panel compares the age distribution of household members in the households screened for the NLSY97 and the PAY97 to the age distribution in the March 1997 CPS. The right panel compares the family income of youth in the NLSY97 to the family income of households with resident youth aged 12 to 16 in the March 1997 CPS. Both figures adapted from Moore et al. (2000). Panel (a) corresponds to Figure 5.5, panel (b) corresponds to Figure G-1.

To understand how the documented selection at the recruitment stage affects panel dynamics observed over subsequent survey rounds, we compare employment rates in the NLSY97 to contemporaneous employment rates in the Current Population Survey's Annual Social and Economic Supplement (CPS ASEC) over time.[57] To this end, we construct comparable measures of employment in both surveys following Bick et al. (2024). In the NLSY97, we construct the annual hours worked of an individual from the usual weekly hours worked in each of their jobs and the number of weeks

[55]Note that an original investigation by the National Opinion Research Center (Moore et al., 2000) did not yield any conclusive results on the origins of differences in age distributions. However, manipulation by data collectors was not explicitly considered.

[56]See Moore et al. (2000), Appendix G for details.

[57]The CPS ASEC is also referred to as the March CPS.

they worked in each of their jobs in the last calendar year.[58] In the CPS ASEC, we combine information on the the number of weeks worked in the last calendar year and the usual weekly hours worked. Based on the annual hours worked, we build two measures of employment: (i) whether an individual has worked any hours, and (ii) whether an individual has worked more than 520 hours. Below we show results for the latter employment definition, but results are very similar for the former.

Our comparison focuses on cohorts born in 1980 and 1981 who were 15 and 16 years old in the first NLSY97 round.[59] We do not include the younger NLSY 97 cohorts aged 12 to 14 in 1997 because employment questions are restricted to ages 15 and above in the CPS. Throughout we use the initial 1997 cross-sectional weights provided by the NLSY79 and the ASEC person weights provided by the CPS.

We find that the dynamics of the gender gap in youth employment differ substantially between the two datasets – see Figure 12. While gaps are initially similar across surveys, they diverge markedly over time, with significant differences emerging after six years and leading to a 12pp difference after ten years.[60]

It is important to note that differences between the NSLY97 and the CPS ASEC arise as the result of three margins of selection into the NLSY97 sample. The first margin is endogenous sample selection at the screening stage. The second margin is non-participation: 9.3% of eligible youth identified in the screening process opted out of the study. The third margin is attrition of study participants over time. We cannot directly separate the impact of these different margins from each other, such that our comparison with the CPS captures the joint impact of all three.

Interestingly, the recruitment process for the National Longitudinal Survey of Youth 1979 (NLSY79) was markedly different from the one for the NLSY97. In fact, it was purposefully designed to minimize the potential for endogenous sample selection.[61] Rather than employing a 'screen-and-go' approach, screening was separated from the first interview round. Moreover, data collectors were not informed about the age groups that would be included in the longitudinal study when they conducted the screening interviews in 1978. Eligible household members were only determined

58 We impose a cap of 98 hours/week, which corresponds to a 14-hour workday for seven days/week.

59 In both the NLSY97 and the CPS ASEC, we calculate the birth year as the survey year minus the age of the respondent at the time of the survey.

60 Interestingly this development is driven by employment gaps between surveys that have opposite signs for men and women, as shown in Figure A25.

61 See the website of National Longitudinal Surveys by the US Bureau of Labor Statistics for further details, in particular on the sample design and screening process for the NLSY79.

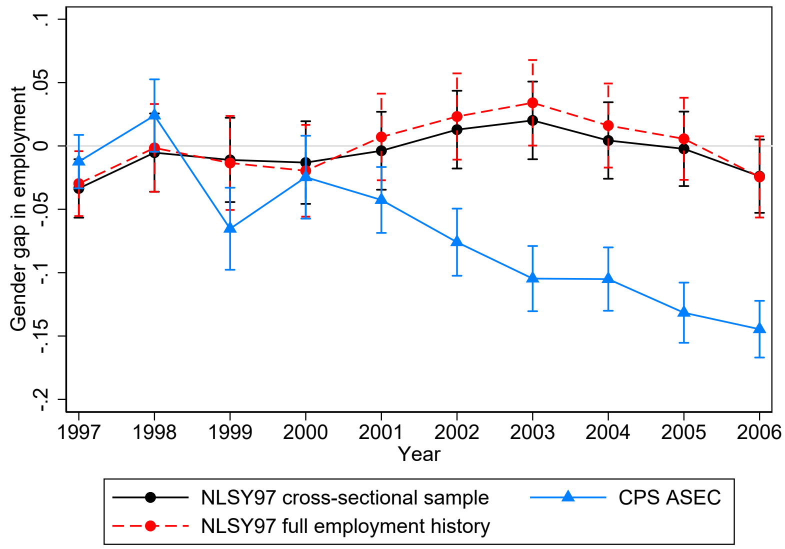


Figure 12: The gender gap in US youth employment: NLSY97 vs. CPS-ASEC

This figure displays estimates of the gender gap in employment for cohorts born in 1980 and 1981 in the NLSY97 and CPS ASEC over time. Individuals are considered employed if they have worked at least 520 hours in the past calendar year. The NLSY97 full employment history sample is restricted to NLSY97 participants with complete weekly employment histories from 1997 to 2006. Markers indicate point estimates, bars 95% confidence intervals.

after screening had been completed and the first round of NLSY79 interviews was conducted in a new field operation in 1979.

In light of the distinction in the recruitment process between the two longitudinal studies, it is noteworthy that we find much smaller differences in employment patterns between initially 15 and 16 year-olds in the NLSY79 and the CPS ASEC (see Figure A26). This lends support to our hypothesis that the screening process for the NLSY97 accounts for a potentially large part of the observed differences in employment dynamics between the NLSY97 and the CPS ASEC.

The differential divergence of employment patterns in the NLSY79 and NLSY97 from patterns in the CPS ASEC has direct implications for research, especially for studies comparing outcomes between both longitudinal datasets (Deming, 2017; Lindenlaub, 2017). It points to differences in participant selection in the two datasets that pose a challenge to the interpretation of differences in outcomes between them.

Finally, the case of the NLSY97 highlights that endogenous sample selection is not limited to low- and middle-income countries. It also affects data collection in high-income countries and has long been a concern among practitioners.[62]

[62] See Judkins et al. (1999) for examples of investigations into determinants of US survey coverage.

7 Discussion

The above evidence from various datasets and contexts points to an underlying trade-off in the collection of data: does more information asked about respondents, ceteris paribus, lead to more endogenous sample selection by data collectors, and more bias?

Before we estimate such an elasticity, two aspects are of note: first, endogenous sample selection is conceptually distinct from non-response bias (Rubin, 1976; Meyer et al., 2015; Dutz et al., 2021). Whereas under non-response bias the econometrician knows that a given subject did not respond and can access at least basic demographic information from the initial listing, endogenous sample selection has the econometrician 'fly blind' – subjects that are screened out are either not recorded at all, or recorded only with manipulated demographics. Second, randomisation of question load provides a way forward: it allows for the estimation of the elasticity of sample size with respect to question load (Subsection 7.1), it opens up an avenue to correcting for selection *ex post* (Subsection 7.2), and it can help understand which innovations in data collection may prevent it *ex ante* (Subsection 7.3).

7.1 Information-bias trade-off in data collection

Does a systematic relationship between question load and endogenous sample selection exist? Our evidence provides a compelling, affirmative answer: across 181 surveys in 73 countries covering more than 30 years of data, we estimate a remarkably stable elasticity of sample size (i.e., missing individuals) with respect to question load.

Descriptively, we observe more eligible men missing in surveys with higher question load in their man's questionnaire relative to the household roster. As reported in Subsection 4.2, the causal elasticity for men, estimated from the random assignment of the man's questionnaire is on average, -0.010, and on average -0.008 for women, estimated from the comparison of surveys with adjacent censuses. These average elasticities across surveys capture the essence of an underlying trade-off that asking for more information (i.e., more questions per respondent) will come, ceteris paribus, at the expense of more bias (i.e., more units being non-randomly excluded).

To put this 'cost of asking (for) more' into context, we provide an illustration from the history of the DHS and MICS programs. Since the 1990s, the length of individual questionnaires in both survey programs proliferated steeply. As Table A11 shows, the average length of the man's questionnaire in our sample nearly doubled, increasing from 103 to 205 questions (columns 1 and 2). At the same time, the elasticity of sample size with respect to question load, if anything, fell (columns 3 and 4). The combination of proliferation in question load and weakly falling elasticity lead to more missing men over time (columns 5 and 6), rising from 6.1% in the 1990s to 8.9% today.

This long-term upward trend in missing individuals is astonishing given simultaneous technological advances for data collection.[63] The historical evidence provides a cautionary tale that the ever-increasing demand for more information by researchers and policymakers comes at substantial cost in terms of endogenous sample selection, and the resulting bias in the collected data.

7.2 Correcting for endogenous sample selection

How could endogenous sample selection be corrected for *ex post* in existing data? Exploiting population censuses from adjacent years allows us to compare marginal distributions of observables in DHS and MICS surveys to the respective distributions in the census to perform one common type of re-weighting, 'raking', to correct for selection on observables. Re-weighted samples display substantial remaining bias in aggregate statistics such as fertility, correcting for about half of the unweighted bias.

Our correction methodology is standard and aims to emulate the situation in which end-users of survey data would find themselves once they suspect the presence of endogenous sample selection. Faced with potentially biased estimates of outcome variables, a correction approach could proceed as follows: find marginal distributions of population parameters for variables collected for every individual in the survey, re-weight observations in the survey to match the population distribution, re-estimate aggregate statistics or regressions using the re-weighted sample.

Commonly called 'raking', we implement this re-weighting procedure by focusing on the subset of survey samples for which survey-census-pairs can be formed to leverage the census to provide marginal distributions of population parameters. We obtain marginal distributions of the maximum number of variables asked in most census and survey pairs, i.e. age, relationship to household head, years of schooling and marriage status.[64] We then rake the survey sample weights using iterative post-stratification until the survey's marginal distributions are jointly indistinguishable from the census'

[63] See Subsection 7.3 below for an explicit test of the effect of different technological innovations.

[64] To account for focal-number bunching of age in censuses and due to the scarcity of the age distribution in some survey samples, we aggregate age into standard five-year bins. Years of schooling is aggregated to four bins: no, primary, secondary or tertiary education.

distribution of the same variables.[65] Finally, we re-estimate our main fertility results using the re-weighted sample. Figure A27 compares original with re-weighted survey estimates for women's number of children ever born compared to the census.

Out of the 34 survey-census pairs that have all listing variables available for raking, 29 pairs are statistically significantly positive in un-weighted specifications. After re-weighting, 15 pairs still remain statistically significantly positive. Before correction, mean bias among those with statistically significant positive bias was 0.12 additional children ever born, whereas correction reduces this to 0.06 for the original 29 pairs and 0.09 additional children ever born for the remaining 15 pairs.[66]

This correction exercise provides two novel insights: first, selection on observables appears to be a major driver of bias in aggregate statistics (here: fertility). Except for three countries, all re-weighted estimates of fertility relative to an adjacent census fall below their unweighted counterparts. Second, although correction reduces bias, we also find strong evidence of remaining bias, suggesting additional selection on unobservables. Data collectors appear to use more information on the ground to identify high effort cost individuals than the few variables they record on the household roster.

Overall, our correction results echo Dutz et al.'s (2021) findings on non-response bias that selection on unobservables presents serious challenges in surveys that are hard to correct for using standard techniques. We leave a Heckman-type correction approach, where the selection function could be recovered from the causal man's questionnaire estimates, for future work.

7.3 Preventing endogenous sample selection

How could endogenous sample selection be prevented *ex ante* in future data collection? Throughout the lifetime of the DHS and MICS, several innovations to improve data collection were introduced. We provide suggestive evidence on the impact of potential remedies such as digital collection tools (e.g., tablets), on-the-ground verification (e.g., field check tables), or credible detection threats (e.g., mandatory audits), and provide brief policy recommendations on design choices that have the potential to limit endogenous sample selection in future data collection.

[65] Results for single-variable raking, when using, for example only individuals' age bin, are qualitatively unchanged, although the bias correction is less effective than multivariate raking.

[66] As proof-of-concept, we also perform re-weighting on the men's fertility sample of surveys with a randomised men's questionnaire, where the control group's marginal distributions of listing variables provide an imperfect proxy of the underlying population. Results are qualitatively unchanged.

To address this question on prevention options, we manually code up details on survey implementation and fieldwork from the official reports accompanying all 181 surveys in our main sample. We focus on three survey features that are systematically documented: the use of (i) mandatory re-interviewing, (ii) field check tables and (iii) tablets.[67] We correlate these features with the estimated elasticities of sampled men. Results are reported in Table A12. We find that mandatory re-interviewing of a fixed fraction of households in each enumeration area is strongly positively correlated with the elasticity, suggesting that this form of auditing significantly reduces manipulation of household rosters by data collectors. The use of field check tables, on the other hand, is not correlated with the elasticity, and the use of tablets is *negatively* correlated with it (after controlling for mandatory re-interviewing), indicating that these features are unlikely to mediate the information-bias trade-off.

Therefore, one policy recommendation for data collection is to re-allocate scarce funding from first-round data collection to second-round audits. A related recommendation (albeit not tested in this paper) would be to break the link between roster inclusion of individuals causing eventual high effort cost by dividing labour between listing and questionnaire tasks. Adversarial incentive schemes, in which two (teams of) data collectors compete over the truthful listing of the same respondent population, represent a promising combination of the previous two recommendations which invites experimentation in future data work.

Finally, one definite policy recommendation arising from this paper is that randomisation in data collection designs provides valuable opportunities for diagnostic analysis – and can help correct for manipulation such as endogenous sample selection.

8 Conclusion

Descriptive statistical analysis and causal inference lie at the heart of empirical research in academia. Causal inference in the social sciences and economics in particular was revolutionised by the introduction of experimental methods in the early 2000s, with identification the subject of much methodological work since. In contrast, data-generating processes have received considerably less attention. However, good data is paramount for both causal inference and descriptive analysis (McKenzie & Rosenzweig, 2012; Dillon et al., 2020).

[67]See Appendix A.1.5 for details.

This paper examines the production of human-collected data, crucial input to a wide range of data sources in the social sciences and beyond. We show that data collectors systematically screen out units that require disproportionate effort in collection based on ex ante-observable characteristics – either by omitting such units entirely or by manipulating their eligibility criteria. This data collector behaviour induces selection of units out of sample, and as a result biases statistics and analysis.

We leverage two complementary empirical strategies: the first strategy exploits random assignment of individual questionnaires, a considerable source of extra work for data collectors, across households in two global household survey programs, the DHS and the MICS. The second strategy compares survey and adjacent census household rosters. We find that in 110 (39) out of 181 surveys with randomised individual questionnaire at least 5% (10%) of eligible subjects are missing from the sample, with an average of 6.5%. Results from the second strategy confirm these results qualitatively and quantitatively: we find an average lower (upper) bound of 6% (12%) missing subjects. Exploiting day-to-day, within-data collector variation in temperature, we confirm that the expected higher effort cost of working under extreme temperatures is indeed reflected in data collectors excluding more household members on hot days.

Endogenous sample selection by data collectors is not random. In our context, missing household members are best described as belonging to more marginalised populations. They are more likely to be peripheral in their respective households in terms of genealogical distance to the household head, more likely to be less educated, poorer, sicker and possibly disabled. This choice of subjects to be screened out is in line with the predictions of a basic theoretical framework that data collectors maxmimise utility by choosing to not report those subjects that are disproportionately more costly (i.e. those eligible for extra questionnaires if included) and subjects who are less likely to be detected as missing by supervisors (i.e. those peripheral to the household with high plausible deniability). Non-random sample selection leads to bias in important aggregate statistics that the DHS and MICS were designed to track closely, such as marriage and fertility.

Three applications showcase our main takeaway that endogenous sample selection is anything but innocuous for empirical research: first, that selection in data collection can interact with supposedly exogenous treatment variation. Second, that it can materially alter economy-wide distributions crucial for estimating macroeconomic frictions and misallocation. Third, that non-random selection at baseline can exacerbate over time since included and excluded units likely lie on different trajectories for dynamic economic outcomes. Accordingly, we document that some empirical phenomena are in fact artefacts of the underlying data collector incentives. For example, the sensitivity of fertility in sub-Saharan Africa to climate shocks such as abnormal droughts or floods is significantly overestimated since not only fertility responds to local climate shocks, but also data collectors effort cost. The 'missing middle' of firm sizes in India most likely represents an artefact of data collector incentives that lead them to artificially depress medium-sized firms' size to reduce the cost of collecting more data for firms above a idiosyncratic size threshold. Therefore, any treatment that increases firms' size across the data collection threshold is likely to be underestimated. Finally, even in longitudinal data collected in high-income country contexts such as the NLSY97, static endogenous sample selection escalates over time, leading to a growing divergence in gender gaps in youth employment compared to datasets without similar data collector incentives.

Our work has implications for human-collected data more broadly, beyond the DHS, MICS, Indian EC and NLSY97. Endogenous sample selection appears to affect other household surveys such as living standard measurement and household budget surveys (see Figure A28), labour force and time-use surveys (see Figure A29), as well as firm and farm censuses. We leave a rigorous analysis of bias in the latter, where sample selection is anecdotally ripe, for future work.

In the grand scheme of things, the design of data collection efforts represents an implicit decision on where the policymaker or researcher intends to land on an underlying possibility frontier that trades off more information against more bias. The more information researchers request in the form of more questions, more modules, more eligible respondents, the more biased the enumerated sample. In the absence of any universal remedy to mute, counter or correct for data collector incentives, this paper shows that policymakers and researchers would benefit from taking this trade-off into account, designing and conducting data collection accordingly.

Finally, whereas the cost of reducing bias in data collection fall squarely on the data collecting entity, e.g. national statistical offices, the benefits of curtailing endogenous sample selection represent a public good. Therefore, we call for more collaboration and exchange between researchers and data collecting entities.

References

Abay, K. A., Berhane, G., Hoddinott, J., & Tafere, K. (2022). Respondent Fatigue Reduces Dietary Diversity Scores Reported from Mobile Phone Surveys in Ethiopia during the COVID-19 Pandemic. *J Nutr.*, *152*(10), 2269–2276.

Abreha, K. G., Cirera, X., Davies, E. A. R., Fattal Jaef, R. N., & Maemir, H. B. (2022). *Deconstructing the Missing Middle : Informality and Growth of Firms in Sub-Saharan Africa* (Policy Research Working Paper Series No. 10233). The World Bank.

Adhvaryuy, A., Kalaz, N., & Nyshadham, N. (2019). The light and the heat: Productivity co-benefits of energy-saving technology. *The Review of Economics and Statistics*, *102*, 1–36.

Ambler, K., Herskowitz, S., & Maredia, M. K. (2021). Are we done yet? Response fatigue and rural livelihoods. *Journal of Development Economics*, *153*, 102736.

Amirapu, A., & Gechter, M. (2020). Labor Regulations and the Cost of Corruption: Evidence from the Indian Firm Size Distribution. *The Review of Economics and Statistics*, *102*(1), 34–48.

Anderson, S. (2018). Legal origins and female hiv. *American Economic Review*, *108*(6), 1407–39.

Asher, S., Campion, A., Gollin, D., & Novosad, P. (2024). *The Long-Run Development Impacts of Agricultural Productivity Gains: Evidence from Irrigation Canals in India* (CEPR Discussion Papers No. 17414). C.E.P.R. Discussion Papers.

Asher, S., & Novosad, P. (2017). Politics and local economic growth: Evidence from india. *American Economic Journal: Applied Economics*, *9*(1), 229–73.

Asher, S., & Novosad, P. (2020). Rural roads and local economic development. *American Economic Review*, *110*(3), 797–823.

Axtell, R. L. (2001). Zipf distribution of u.s. firm sizes. *Science*, *293*(5536), 1818–1820.

Bardasi, E., Beegle, K., Dillon, A., & Serneels, P. (2011). Do Labor Statistics Depend on How and to Whom the Questions Are Asked? *World Bank Economic Review*, *25*(3), 418–447.

Becker, A. (2024). On the economic origins of concerns over women's chastity. *Review of Economic Studies*, rdae084.

Bick, A., Blandin, A., & Rogerson, R. (2024). *After 40 years, how representative are labor market outcomes in the nlsy79?* (Working Papers No. 2024-10). Federal Reserve Bank of St. Louis.

Blakeslee, D., Dar, A., Fishman, R., Malik, S., Pellegrina, H. S., & Bagavathinathan, K. S. (2023). Irrigation and the spatial pattern of local economic development in india. *Journal of Development Economics*, *161*, 102997.

Blasius, J., & Thiessen, V. (2015). Should we trust survey data? assessing response simplification and data fabrication. *Social science research*, *52*, 479–93.

Buera, F. J., Kaboski, J. P., & Shin, Y. (2020). The macroeconomics of microfinance. *The Review of Economic Studies*, *88*(1), 126–161.

Burke, M., Gong, E., & Jones, K. (2015). Income shocks and hiv in africa. *The Economic Journal*, *125*(585), 1157–1189.

Burlig, F., & Preonas, L. (2024). Out of the Darkness and into the Light? Development Effects of Rural Electrification. *Journal of Political Economy*, *132*(9), 2937–2971.

Cai, X., Lu, Y., & Wang, J. (2018). The impact of temperature on manufacturing worker productivity: Evidence from personnel data. *Journal of Comparative Economics*, *46*.

Carleton, T. A., & Hsiang, S. M. (2016). Social and economic impacts of climate. *Science*, *353*(6304), aad9837.

Castorena, O., Cohen, M. J., Lupu, N., & Zechmeister, E. J. (2023). How Worried Should We Be? The Implications of Fabricated Survey Data for Political Science. *International Journal of Public Opinion Research*, *35*(2), edad007.

Chatterjee, S., & Vogl, T. (2018). Escaping Malthus: Economic Growth and Fertility Change in the Developing World. *American Economic Review*, *108*(6), 1440–67.

Chiplunkar, G., & Goldberg, P. K. (2024). Aggregate implications of barriers to female entrepreneurship. *Econometrica*, *92*(6), 1801–1835.

Corno, L., Hildebrandt, N., & Voena, A. (2020). Age of marriage, weather shocks, and the direction of marriage payments. *Econometrica*, *88*(3), 879–915.

Corno, L., & Voena, A. (2023). Child marriage as informal insurance: Empirical evidence and policy simulations. *Journal of Development Economics*, *162*, 103047.

Couttenier, M., & Soubeyran, R. (2014). Drought and civil war in sub-saharan africa. *The Economic Journal*, *124*(575), 201–244.

Crespi, L. P. (1945). The Cheater Problem in Polling. *Public Opinion Quarterly*, *9*(4), 431–445.

Cui, W., Cao, G., Park, J., Ouyang, Q., & Zhu, Y. (2013). Influence of indoor air temperature on human thermal comfort, motivation and performance. *Building and Environment*, *68*, 114–122.

Deming, D. J. (2017). The growing importance of social skills in the labor market*. *Q J Econ*, *132*(4), 1593–1640.

Dervisevic, E., & Goldstein, M. (2023). He said, she said: The impact of gender and marriage perceptions on self and proxy reporting of labor. *Journal of Development Economics*, *161*, 103028.

Di Maio, M., & Fiala, N. (2019). Be Wary of Those Who Ask: A Randomized Experiment on the Size and Determinants of the Enumerator Effect. *The World Bank Economic Review*, *34*(3), 654–669.

Dillon, A., Bardasi, E., Beegle, K., & Serneels, P. (2012). Explaining variation in child labor statistics [Symposium on Measurement and Survey Design]. *Journal of Development Economics*, *98*(1), 136–147.

Dillon, A., Karlan, D., Udry, C., & Zinman, J. (2020). Good identification, meet good data. *World Development*, *127*, 104796.

Dillon, A., & Mensah, E. (2024). Respondent biases in agricultural household surveys. *Journal of Development Economics*, *166*, 103198.

Dimitrova, A. (2021). Seasonal droughts and the risk of childhood undernutrition in ethiopia. *World Development*, *141*, 105417.

Dupas, P., Falezan, C., Mabeu, M. C., & Rossi, P. (2023). *Long-run impacts of forced labor migration on fertility behaviors: Evidence from colonial west africa* (Working Paper No. 31993).

Dutz, D., Huitfeldt, I., Lacouture, S., Mogstad, M., Torgovitsky, A., & van Dijk, W. (2021). *Selection in Surveys: Using Randomized Incentives to Detect and Account for Nonresponse Bias* (Working Paper No. 29549). National Bureau of Economic Research.

Eckman, S., & Koch, A. (2019). Interviewer involvement in sample selection shapes the relationship between response rates and data quality. *Public opinion quarterly*, *83*, 313–337.

Epstein, A., Bendavid, E., Nash, D., Charlebois, E. D., & Weiser, S. D. (2020). Drought and intimate partner violence towards women in 19 countries in sub-saharan africa during 2011-2018: A population-based study. *PLoS Medicine*, *17*(3), e1003064.

Fichera, E., & Savage, D. (2015). Income and health in tanzania. an instrumental variable approach. *World development*, *66*, 500–515.

Finn, A., & Ranchhod, V. (2015). Genuine Fakes: The Prevalence and Implications of Data Fabrication in a Large South African Survey. *The World Bank Economic Review*, *31*(1), 129–157.

Geruso, M., & Spears, D. (2018). *Heat, humidity, and infant mortality in the developing world* (tech. rep. No. 24870). National Bureau of Economic Research, Inc.

Government of the Republic of Malawi. (2018). *National health policy "towards universal health coverage"* (tech. rep.).

Guarnieri, E., & Tur-Prats, A. (2023). Cultural distance and conflict-related sexual violence. *The Quarterly Journal of Economics*, *138*(3), 1817–1861.

Hernández-Pérez, R., Angulo-Brown, F., & Tun, D. (2006). Company size distribution for developing countries. *Physica A: Statistical Mechanics and its Applications*, *359*, 607–618.

Hjort, J., & Poulsen, J. (2019). The arrival of fast internet and employment in africa. *American Economic Review*, *109*(3), 1032–79.

Horrigan, M., Moore, W., Pedlow, S., & Wolter, K. (1999). Undercoverage in a large national screening survey for youths. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 570–75.

Hsieh, C.-T., & Olken, B. A. (2014). The missing "missing middle". *Journal of Economic Perspectives*, *28*(3), 89–108.

ICF. (1982-2022). *Demographic and health surveys (various)* [Datasets used: Various]. Rockville, Maryland, ICF.

Jayachandran, S., & Pande, R. (2017). Why are indian children so short? the role of birth order and son preference. *American Economic Review*, *107*(9), 2600–2629.

Jeong, D., Aggarwal, S., Robinson, J., Kumar, N., Spearot, A., & Park, D. S. (2023). Exhaustive or exhausting? Evidence on respondent fatigue in long surveys. *Journal of Development Economics*, *161*, 102992.

Judkins, D., DiGaetano, R., Chu, A., & Shapiro, G. (1999). Coverage in screening surveys at westat. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 581–86.

Kilic, T., Moylan, H., & Koolwal, G. (2021). Getting the (Gender-Disaggregated) lay of the land: Impact of survey respondent selection on measuring land ownership and rights. *World Development*, *146*(100).

Kosyakova, Y., Skopek, J., & Eckman, S. (2014). Do interviewers manipulate responses to filter questions? evidence from a multilevel approach. *International Journal of Public Opinion Research*, *27*(3), 417–431.

Le, K., & Nguyen, M. (2022). Droughts and child health in bangladesh. *Plos one*, *17*(3), e0265617.

Lindenlaub, I. (2017). Sorting multidimensional types: Theory and application. *Rev Econ Stud*, *84*(2), 718–789.

LoPalo, M. (2023). Temperature, Worker Productivity, and Adaptation: Evidence from Survey Data Production. *American Economic Journal: Applied Economics*, *15*(1), 192–229.

Lowes, S., & Montero, E. (2021). The legacy of colonial medicine in central africa. *American Economic Review*, *111*(4), 1284–1314.

Marckwardt, A. M., & Rutstein, S. O. (1996). *Accuracy of dhs-ii demographic data: Gains and losses in comparison with earlier surveys* (tech. rep.). Calverton, Maryland, USA, Macro International.

Masselus, L., & Fiala, N. (2024). Whom to ask? Testing respondent effects in household surveys. *Journal of Development Economics*, *168*, 103265.

McKenzie, D., & Rosenzweig, M. (2012). Preface for symposium on measurement and survey design.

Menold, N., Landrock, U., Winker, P., Pellner, N., & Kemper, C. J. (2018). The impact of payment and respondents' participation on interviewers' accuracy in face-to-face surveys: Investigations from a field experiment. *Field Methods*, *30*(4), 295–311.

Meyer, B. D., Mok, W. K. C., & Sullivan, J. X. (2015). Household Surveys in Crisis. *Journal of Economic Perspectives*, *29*(4), 199–226.

Michalopoulos, S., & Papaioannou, E. (2016). The long-run effects of the scramble for africa. *American Economic Review*, *106*(7), 1802–48.

Miguel, E., Satyanath, S., & Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of political Economy*, *112*(4), 725–753.

Ministère de la Santé et de l'Hygiène Publique. (2021). *Plan d'action national budgétisé de planification familiale 2021-2025 de côte d'ivoire* (tech. rep.). République de Côte d'Ivoire.

Ministry of Health, Republic of Ghana. (2020). *National health policy: Ensuring healthy lives for all* (tech. rep.).

Ministry of Health, Republic of Kenya. (2022). *The national reproductive health policy 2022-2032* (tech. rep.).

Ministry of Health, Uganda. (2017). *The national policy guidelines and service standards for sexual and reproductive health and rights* (tech. rep.).

Ministry of Statistics and Programme Implementation, India. (2005). *Instruction Manual Fifth Economic Census (2005).*

Moore, W., Pedlow, S., Krishnamurty, P., Wolter, K., & Chicago, I. (2000). National longitudinal survey of youth 1997 (nlsy97). *National Opinion Research Center, Chicago, IL*, *254*.

Muralidharan, K., Niehaus, P., & Sukhtankar, S. (2023). General equilibrium effects of (improving) public employment programs: Experimental evidence from india. *Econometrica*, *91*(4), 1261–1295.

Nagata, J. M., Epstein, A., Ganson, K. T., Benmarhnia, T., & Weiser, S. D. (2021). Drought and child vaccination coverage in 22 countries in sub-saharan africa: A retrospective analysis of national survey data from 2011 to 2019. *PLoS medicine*, *18*(9), e1003678.

Nolan, L., Lucas, R., Choi, Y., Short Fabic, M., & Adetunji, J. (2017). The contribution of demographic and health survey data to population and health policymaking: Evidence from three developing countries. *Etude de la population africaine*, *31*, 3394–3407.

Norling, J. (2022). Fertility following natural disasters and epidemics in africa. *The World Bank Economic Review*, *36*(4), 955–971.

Paxson, C. H. (1992). Using weather variability to estimate the response of savings to transitory income in thailand. *The American Economic Review*, 15–33.

Pilcher, J. J., Nadler, E., & Busch, C. (2002). Effects of hot and cold temperature exposure on performance: A meta-analytic review. *Ergonomics*, *45*, 682–98.

Pullum, T. W. (2006). *An assessment of age and date reporting in the dhs surveys, 1985-2003* (tech. rep.). Calverton, Maryland, USA, Macro International.

Robbins, M., Johnson, T. P., Pennell, B. E., Stoop, I. A. L., & Dorer, B. (2018). New frontiers in detecting data fabrication. *Advances in Comparative Survey Methods: Multicultural, Multinational and Multiregional Contexts (MC), edited by TP Johnson, B.-E. Pennell, IAL Stoop, and B. Dorer,-. Hoboken, NJ: Wiley*.

Rossi, P. (2018). Strategic Choices in Polygamous Households: Theory and Evidence from Senegal. *The Review of Economic Studies*, *86*(3), 1332–1370.

Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, *63*(3), 581–592.

Ruggles, S., Cleveland, L., Lovaton, R., Sarkar, S., Sobek, M., Burk, D., Ehrlich, D., Heimann, Q., & Lee, J. (2024). *Integrated public use microdata series, international: Version 7.5*. Minneapolis, MN, IPUMS.

Sen, R. (2024). *Supervision at work: Evidence from a field experiment* (tech. rep.). Working paper.

Seppanen, O., Fisk, W., & Lei, Q. H. (2005). Effect of temperature on task performance in office environment.

Serneels, P., Beegle, K., & Dillon, A. (2017). Do returns to education depend on how and whom you ask? *Economics of Education Review*, *60*, 5–19.

Short Fabic, M., Choi, Y., & Bird, S. (2012). A systematic review of demographic and health surveys: Data availability and utilization for research. *Bulletin of the World Health Organization*, *90*, 604–12.

Sustainable Development Solutions Network. (2015). Data for development: A needs assessment for sdg monitoring and statistical capacity development [Paris, France and New York, USA: SDSN].

Thiede, B. C. (2014). Rainfall shocks and within-community wealth inequality: Evidence from rural ethiopia. *World Development*, *64*, 181–193.

Tybout, J. (2000). Manufacturing firms in developing countries: How well do they do, and why? *Journal of Economic Literature*, *38*(1), 11–44.

Tybout, J. (2014). The missing middle: Correspondence. *Journal of Economic Perspectives*, *28*(4), 235–36.

UNICEF. (2000-2022). *Multiple indicator cluster surveys (mics)* [Datasets used: Various]. New York, USA, UNICEF [Distributor].

Vogl, T. S. (2013). Marriage Institutions and Sibling Competition: Evidence from South Asia. *The Quarterly Journal of Economics*, *128*(3), 1017–1072.

Vogl, T. S. (2016). Differential Fertility, Human Capital, and Development. *The Review of Economic Studies*, *83*(1), 365–401.

West, B., & Blom, A. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, *5*, 175–211.

Young, A. (2012). The african growth miracle. *Journal of Political Economy*, *120*(4), 696–739.

Zipfel, C. (2024). *The demand side of africa's demographictransition: Desired fertility, wealth, and jobs* (tech. rep.). Stockholm School of Economics.

A Online appendix

A.1 Data

A.1.1 Selection of surveys

The main criterion for the inclusion of a survey into our main sample is the administration of a man's questionnaire in a randomly selected subset of households. Additionally, we restrict our sample to nationally representative surveys. This enables us to examine implications of endogenous sample selection for national statistics.

We identify relevant surveys from the official survey reports published on the DHS and MICS websites. To this end, we read more than 800 reports in five different languages and extract information on all survey components that were randomly varied across households, most importantly the man's questionnaire, biomarker collection and the domestic violence module. The combination of the information from the reports and the microdata allows us to understand the underlying randomisation in detail. In particular, we pay close attention to the manner in which different randomised survey features were either bundled or cross-randomised and the respective treatment probabilities.

Among all 236 surveys that satisfy our criteria, we exclude 55 because they do not lend themselves to our analysis due to differences in survey design or data issues. All excluded surveys and the respective reasons for exclusion are listed in Table A1. First, we exclude 28 surveys that administered additional survey features, such as biomarker collection among children, in control households (without a man's questionnaire) that were not implemented treatment households. In these cases, differences in outcomes between treatment and control households cannot be attributed solely to the man's questionnaire. Second, we exclude 13 surveys in which eligibility for the man's questionnaire is conditional on marital status. Selection into individual questionnaires in these surveys is not comparable to selection in included surveys and thus results would not be directly comparable. Moreover, the resulting samples are not nationally representative. Third, we exclude 9 MICS due to data issues. For 6 MICS in which sampling is stratified by enumeration area and the presence of children in the household, we do not observe the latter stratification variable in the microdata. Thus we cannot control for stratum fixed effects. For 3 MICS, we are not able to merge the individual- and household-level microdata source files because identifiers do not match across files. Fourth, 3 DHS are excluded because their man's questionnaire does not have an upper age limit, thereby not allowing us to define a comparable group of ineligible men in these surveys. Finally, one DHS is excluded because treatment was randomised across enumeration areas rather than across households within enumeration areas, making comparisons with other surveys difficult, and one MICS is excluded due to contradicting information about treatment assignment in the survey report and the microdata.

A.1.2 Eligibility for individual questionnaires

To determine the age thresholds for the eligibility of household members for individual questionnaires, we systematically extract information on the age thresholds from the official survey reports and questionnaires for all surveys in our sample. Subsequently, we verify the consistency of the microdata with these thresholds.

A.1.3 Data collector effort cost

We construct two proxies of the effort cost associated with household members of a given sex and age.

Questions listed. For all surveys in our respective samples, we count the total number of questions contained in the household roster (individual-level questions in the household questionnaire), the man's questionnaire and the woman's questionnaire. We proxy the effort cost associated with a (wo)man of eligible age with the sum of the number of questions in the roster and the individual questionnaire. The effort cost associated with ineligible household members is measured by the number of questions in the household roster.

We count questions as follows. We follow the numbering of questions in the official questionnaires and do not count sub-questions. For example, questions 32, 32A and 32B are counted as single question. Note that a small set of questions may be repeated multiple times for the same respondent. For example, women in recent DHS are asked several questions about each birth they have ever given. Independent of the number of births a woman has given, we only count each of these questions once. To ensure accurate counting, we conduct two independent counts for a sub-sample of 33 surveys. Reassuringly, we find a correlation coefficient of above 0.99 between counts, with a mean absolute deviation of less than 1%.

When counting questions in population and housing censuses, we differentiate between individual-level questions asked to women of fertile age (typically 12 years and older) and all other individual level questions. We think of the former questions as the equivalent of the woman's questionnaire and the latter questions as the equivalent of the household roster in the DHS and the MICS.

Questions asked. The number of questions asked to a given respondent is usually smaller than the total number of questions contained in questionnaires. This is because certain subsets of questions are only asked to respondents with specific characteristics. For example, in the MICS only women of eligible age who have ever given birth are asked about their birth history. To count the number of questions actually answered by each respondent, we manually match each question in the questionnaire with the corresponding variable in the microdata. In the MICS, there is a one-to-one link between questions listed in the questionnaire and variables in the dataset. Moreover, variable names in the microdata follow the question numbering in questionnaire, facilitating the matching. In the DHS and the PHC, this is not the case. IPUMS source variables have descriptive variable names that help with matching. DHS matching relies on variable labels and tabulations as variable names cannot be used due to DHS recoding process that names variables using standardized codes (e.g., hv104). Given the large number of questions in the DHS, the resulting matching process is very tedious and time-consuming (5-8 hours per survey). Therefore, we only conduct this exercise for a subset of DHS (31) while we complete it for all MICS in our sample.

In each of the three data sources, we ensure a variable is coded as missing if and only if the matched question was not asked about a given individual. Subsequently, we count the number of non-missing entries across all variables for each household member. To obtain the a measure of the effort cost associated with a given sex and age, we average the number of questions asked within sex-age cells.

A.1.4 Outcome variables

Ever married. We define having ever been married in a broad sense. In line with most surveys in our sample, we count all individuals that are married, living with a partner, separated, divorced or widowed as ever married. Information on the marital status is collected through different questionnaires in the surveys we work with. In the MICS, marital status is asked in the individual questionnaire, not in the household roster. The DHS initially operated in the same way, but gradually moved to systematically including a question about marital status in the household roster. While the roster only features a question on marital status in a some of the DHS conducted prior to 2012, it includes such a question for all surveys in our sample conducted thereafter. So, we observe the marital status of men in control households in all DHS conducted post 2012 and a subset of DHS conducted earlier.

Close relationship to household head. Nearly all censuses and surveys in our samples elicit information on the relationship of household members to the household head. The set of answer options varies greatly across surveys and censuses, however. To harmonise the information, we create an indicator variable that equals to 1 if a household member is closely related to the head of the household and zero otherwise. We define children, spouse(s), parents, parents-in-law and grandchildren as closely related to the head, and other relatives (e.g., uncles) and unrelated household members (e.g., domestic workers) as distantly related.

Years of schooling. Information on years of schooling is readily available in harmonised form in DHS and IPUMS-International census data. In the MICS and non-IPUMS censuses, we harmonise this information ourselves, combining information on the highest level and grade of education completed with the structure of the education system at the time of the survey. Note that we only consider formal education when doing so.

Number of biological children in the household. Most surveys in our sample include a module on the survival of parents in the household roster. For all children aged 17 and below, this module asks whether the biological mother and father are alive, and if so whether they live in the household. If the answer to both of these questions is affirmative, their line number is recorded. We measure the number of biological children each household member lives with by counting the number of children in the household for which they are indicated as the parent.

Children ever born. This variable captures the total number of children ever born alive to a woman. It is top-coded in some population censuses. To ensure comparability with matched surveys, we apply the same top-coding to the matched surveys.

Poor. This variable is an indicator variable that takes value one if an individual does not possess all of the following three items: shoes, clothes and a blanket. Otherwise, it takes the value zero. The underlying information is elicited as part of a module on the basic needs of children between the ages of 5 and 17. The module is included in the following DHS in our sample: NGA 2008, HTI 2005, MWI 2010, NAM 2006, UGA 2006.

Sick. This variable is an indicator variable that takes value one if an individual has been very sick for at least 3 months during the past 12 months. Otherwise, it takes the value zero. The underlying information is elicited as part of a module on chronic disease which is limited to adults between the ages of 18 and 59 in most surveys. The module is included in the following DHS in our sample: MLI 2006, NER 2006, RWA 2005, SEN 2005, UGA 2006, NGA 2008, HTI 2005, COD 2007 and MWI 2004.

Disabled. This variable is an indicator variable that takes value one if an individual is classified as disabled and zero otherwise. It is available in the household roster of the following DHS in our sample: BOL 1998, GHA 1998, GMB 2013, KHM 2014, MLI 2018, TZA 2022, UGA 2016 and ZAF 2016. We define a person as disabled if they suffer from at least one form of disability (blind, deaf, etc.). In surveys where the extent to which individuals have difficulties with certain activities (e.g., seeing, hearing, moving) is elicited, we consider individuals as disabled if they cannot do at least one activity at all or they can only do it with a lot of difficulty.

Orphan. This variable is an indicator variable that takes value one if at least one parent of an individual is not alive. Otherwise, it takes the value zero. The underlying information is elicited for children aged 17 and younger in 84 of the DHS in our sample.

A.1.5 Survey characteristics

Reading through the final reports from all 181 surveys in our main sample, we extract information on survey implementation and data processing. We systematically code up the below variables.

Field check tables. We determine if field check tables were used during survey implementation. These tables are sometimes also referred to as quality control tables and contain descriptive statistics of key indicators. They are produced regularly throughout the fieldwork period and are used to provide feedback to supervisors and surveyors.

Mandatory re-interviewing. We identify surveys that conduct mandatory re-interviewing. In this case, typically two sets of households are re-interviewed: first, a random subset of households in each enumeration area and second, all households which have been identified as outliers along key survey dimensions.

Use of tablets. We differentiate between surveys that use paper and tablet questionnaires. In the former case, responses are recorded on paper and later entered into computers. In the latter case, responses are directly recorded on tablets and later transmitted to a central server.

A.2 Mechanisms

A.2.1 Data collector selection

The eligibility of a given household for the man's questionnaire is revealed on the first page of the household questionnaire. In response to this information, supervisors can strategically assign data collectors to households with and without a man's questionnaire. This raises the question how the eligibility of a household for the man's questionnaire affects the identity of the data collector recording the household roster. Leveraging information on the characteristics of data collectors from the DHS fieldworker questionnaire, available for 19 surveys in our sample, we empirically test how data collector characteristics differ between households with and without a man's questionnaire.[68] We find that in most surveys, data collectors in charge of the household roster are significantly less likely to be female in treatment households. The tendency to assign male data collectors to households with a man's questionnaire can be attributed to the survey program's objective to conduct same-sex individual interviews, i.e., to have male data collectors administer man's questionnaires and female data collectors administer woman's questionnaires. This implies that a male data collector is required at households that are eligible for the man's questionnaire, but

[68]The DHS fieldworker questionnaire was introduced in 2015. Hence, data collector information is not available for earlier surveys. The MICS does not publish any data collector characteristics.

not at ineligible households. The effect of the man's questionnaire on age and education varies across surveys, both in sign and magnitude. Experience with previous DHS is negatively affected in most surveys, but also heavily positively affected in a few surveys. Figure A7 displays all the estimates.

In the face of these changes in data collector characteristics, it is important to note that, consistent with the idea of moral hazard, selection of data collectors cannot explain the reductions in the number of eligible men as point estimates are barely affected by the inclusion of data collector fixed effects (see Figure A8).

A.2.2 Respondent selection

The assignment to the man's questionnaire may alter the identity of the respondent to the household roster. In fact, we find that in almost all surveys, respondents in households with a man's questionnaire are less likely to be female, more likely be the household head as well as somewhat older and more educated (see Figure A30).

A.3 Lower and upper bound of missing women

We use the following regression specification to estimate the difference-in-differences of interest:

$$Y_{is} = \beta_0 + \beta_1 SVY_i + \beta_2 Eligible_s + \beta_3 (SVY_i \times Eligible_s) + \mu_{is} \tag{6}$$

where Y_{is} is the number of women of eligibility status $s \in \{eligible, ineligible\}$ recorded in household i. Women are considered eligible if they are in the age range that is eligible for the DHS/MICS woman's questionnaire (usually 15 to 49). They are considered ineligible if they are outside this age range and older than 9 years of age. The lower bound of 9 limits the conflation of the impact of the woman's questionnaire with the impact of the high question load for children under 5 in the DHS/MICS on the presence of ineligible women.[69] SVY_i is an indicator that takes the value one if the household roster was recorded by the DHS/MICS and zero if it was recorded by the census. $Eligible_s$ is an indicator that takes value one if the outcome is the number of eligible household members, and zero if it is the number of ineligible household members. We scale survey sampling weights such that the total number of households in surveys and contemporaneous censuses is identical, and cluster standard errors at the household level. β_3 captures the difference-in-differences of interest. Accordingly, the upper bound of missing women is equal to β_3 and the lower bound is equal to $\beta_3/2$.

[69]This assumes that the high question load for children under 5 may lead to the displacement of their age to values above 5, but rarely above 9.

A.4 Use of DHS and MICS in top demography journals

To determine the number of top demography papers with reference to the DHS or the MICS, we run a search for the keywords "Demographic and Health Survey(s)" and "Multiple Indicator Cluster Survey(s)" across all fields on JSTOR. We implement the search using Constellate, a web-based text analytics service provided by ITHAKA, which allows us to automatically query the JSTOR collection. We restrict the publication time to 2013 to 2017 for data availability reasons and focus on the top two journals in Demography, *Demography* and the *Population and Development Review*.

A.5 Appendix Figures

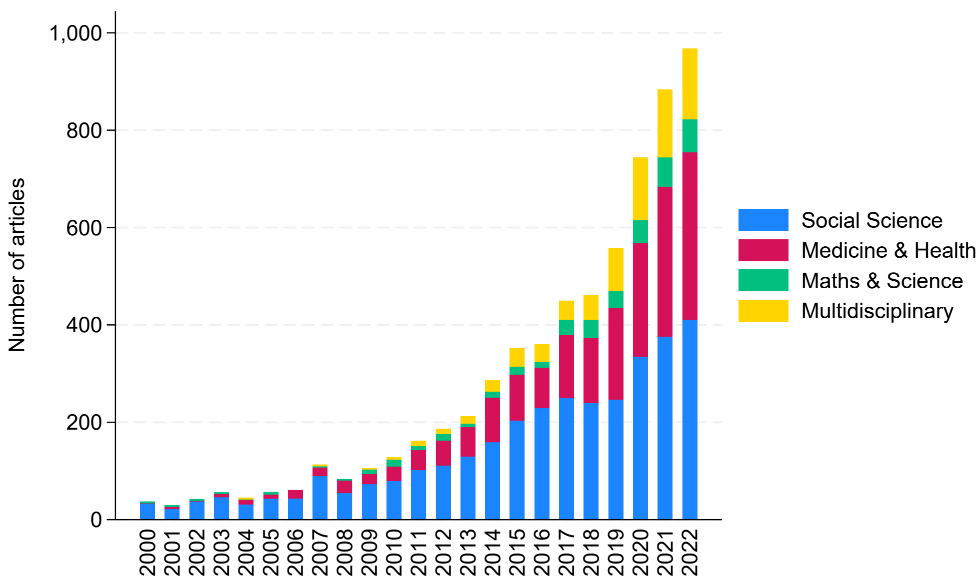


Figure A1: Use of DHS and MICS in research over time

This figure displays the count of journal articles that refer to the DHS or MICS in their title or abstract across disciplines over time. Counts were generated from the Web of Science database using keyword search. The set of journals is restricted to those that formed part of the Essential Science Indicator journal master list as of June 2024.

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 A2: 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 A3: MICS, Ghana 2011: Household roster

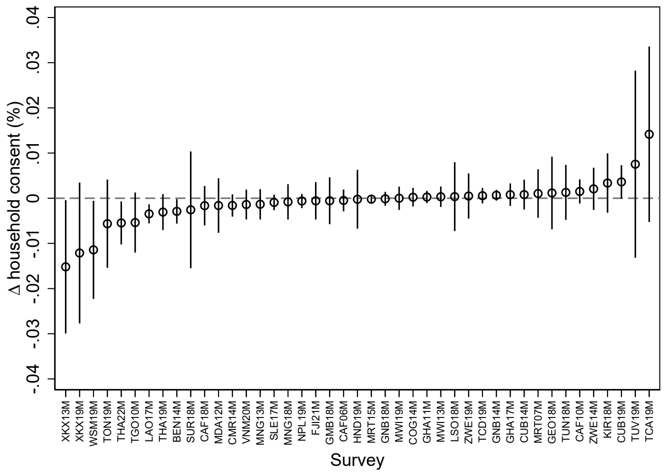


Figure A4: Effect of man's questionnaire on household consent in the MICS

This figure displays estimates of the effect of household assignment to the man's questionnaire on household consent across MICS with a randomly assigned man's questionnaire. Circles indicate point estimates and bars indicate 95% confidence intervals. Surveys are sorted along the x-axis in ascending order of the point estimate. Every survey is labeled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively.

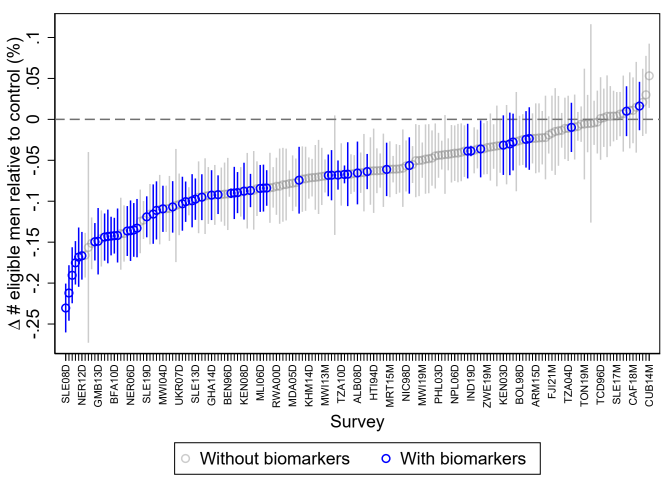


Figure A5: Effect of man's questionnaire with and without male biomarker collection

This figure displays estimates of β from equation (3) relative to the control mean where the outcome variable is the number of eligible men in the household. The sample consists of all 181 DHS and MICS with a man's questionnaire that is randomly assigned across households. Circles indicate point estimates and bars indicate 95% confidence intervals. Estimates from surveys that include biomarker collection from eligible men are shown in blue. Surveys are sorted along the x-axis in ascending order of the point estimate. Every 5th survey is labelled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively. All estimates are reported in Table A4, column (3).

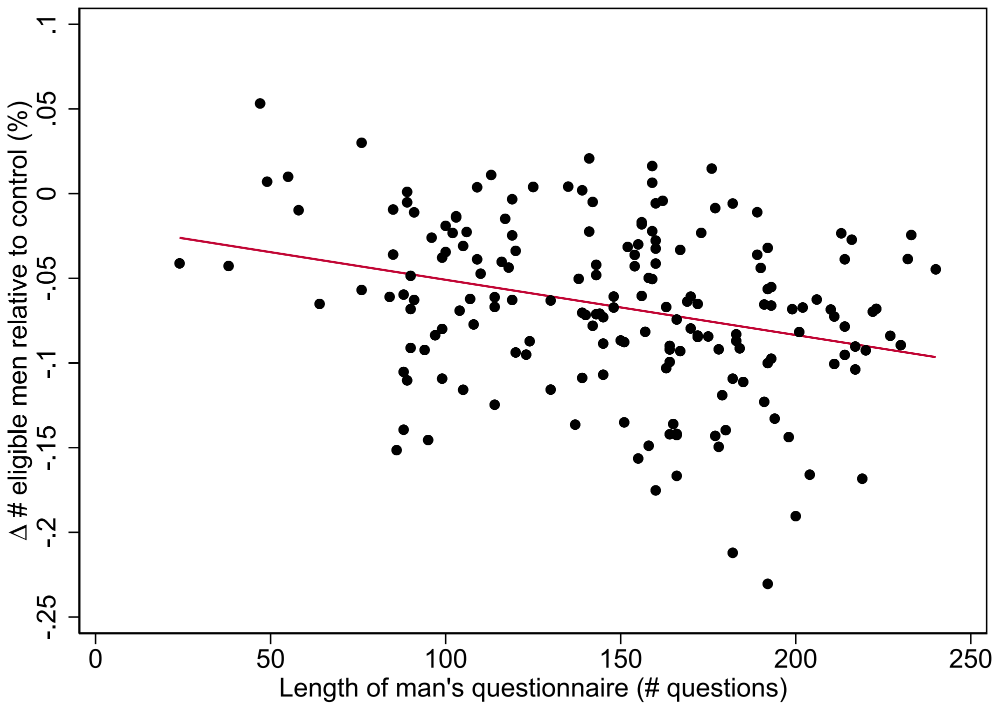
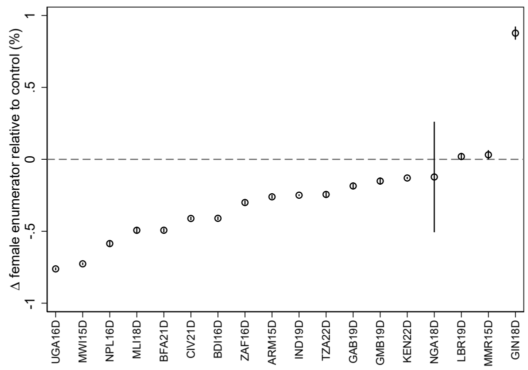
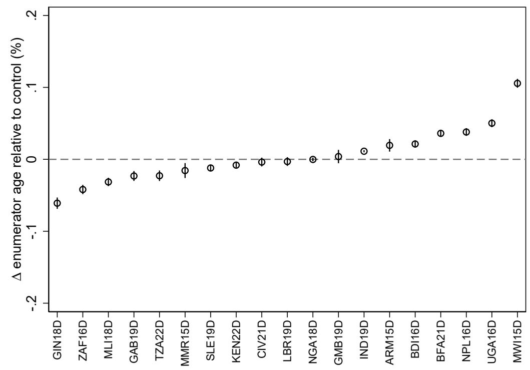


Figure A6: Effect of man's questionnaire vs. length of questionnaire across surveys

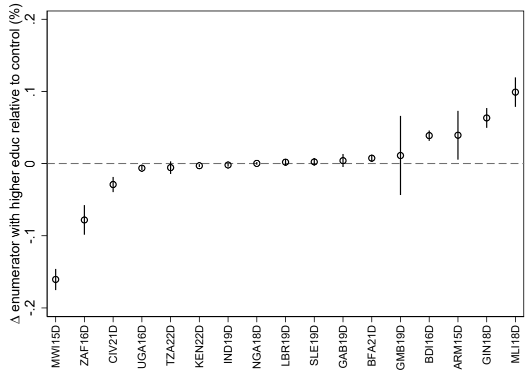
This figure plots estimates of β from equation (3) relative to the control mean against the length of the man's questionnaire across surveys. The sample consists of all 181 DHS and MICS with a man's questionnaire that is randomly assigned across households. See appendix A.1.3 for details on the measurement of the length of questionnaires. The solid line presents a linear fit.



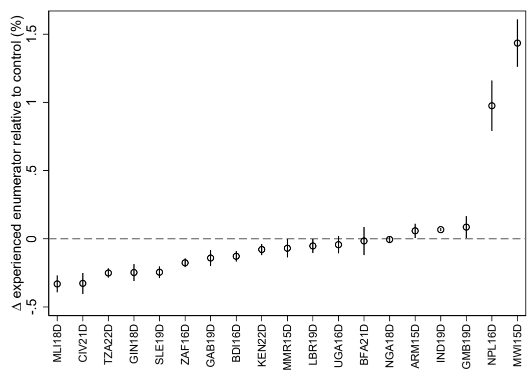
(a) Sex



(b) Age



(c) Higher (post-secondary) education



(d) Previous experience with DHS

Figure A7: Effect of man's questionnaire on data collector characteristics

This figure displays estimates of the effect of household assignment to the man's questionnaire on the characteristics of the data collectors administering the household roster relative to the control group. Standard errors are clustered at the household-level. Circles indicate point estimates and bars indicate 95% confidence intervals. Surveys are sorted along the x-axis in ascending order of the respective point estimate. All surveys are labeled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively.

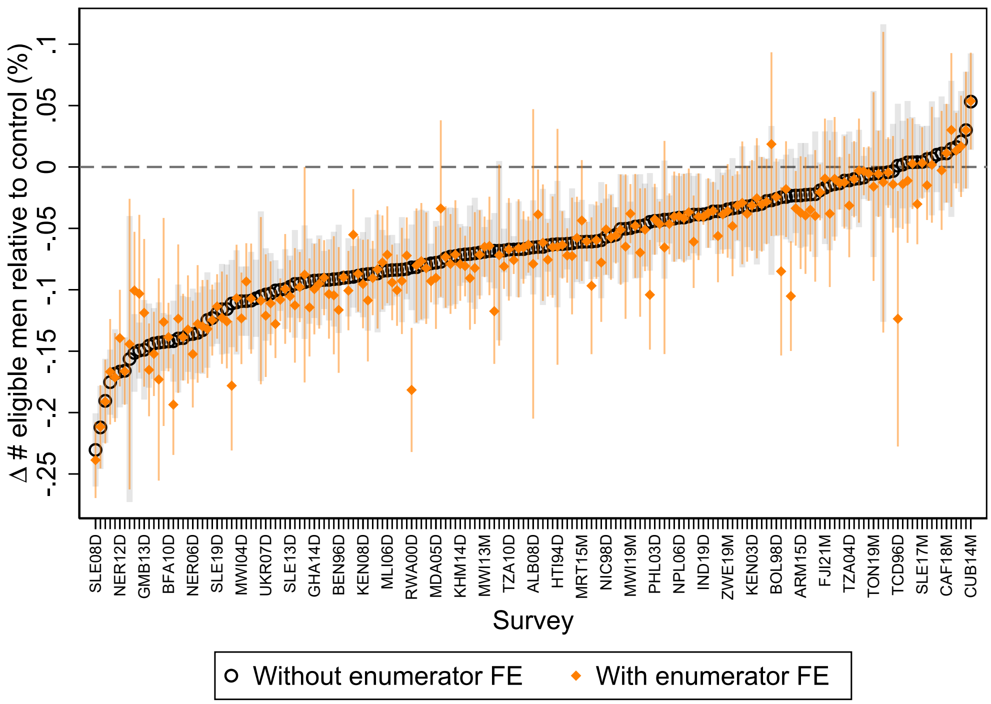


Figure A8: Within-data collector effect of man's questionnaire on # of eligible men

This figure displays estimates of β from equation (3) relative to the control mean where the outcome variable is the number of eligible men in the household. The sample consists of all 181 DHS and MICS with a man's questionnaire that is randomly assigned across households. Circles indicate point estimates and bars indicate 95% confidence intervals. Surveys are sorted along the x-axis in ascending order of the point estimate excluding data collector fixed effects. Every 5th survey is labelled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively.

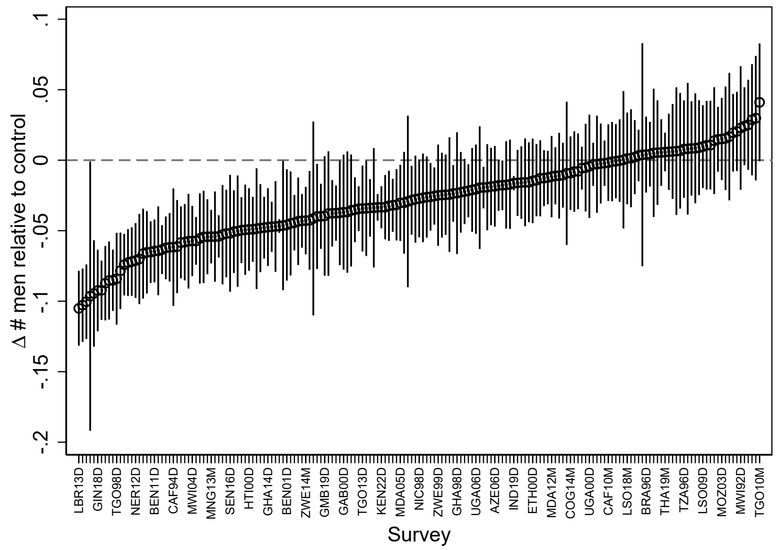


Figure A9: Effect of man's questionnaire on total number of men in the household

This figure displays estimates of β from equation (3) relative to the control mean where the outcome variable is the total number of men in the household. The sample consists of all 181 DHS and MICS with a man's questionnaire that is randomly assigned across households. Circles indicate point estimates and bars indicate 95% confidence intervals. Surveys are sorted along the x-axis in ascending order of the point estimate. Every 5th survey is labelled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively. All estimates are reported in Table A4, column (6).

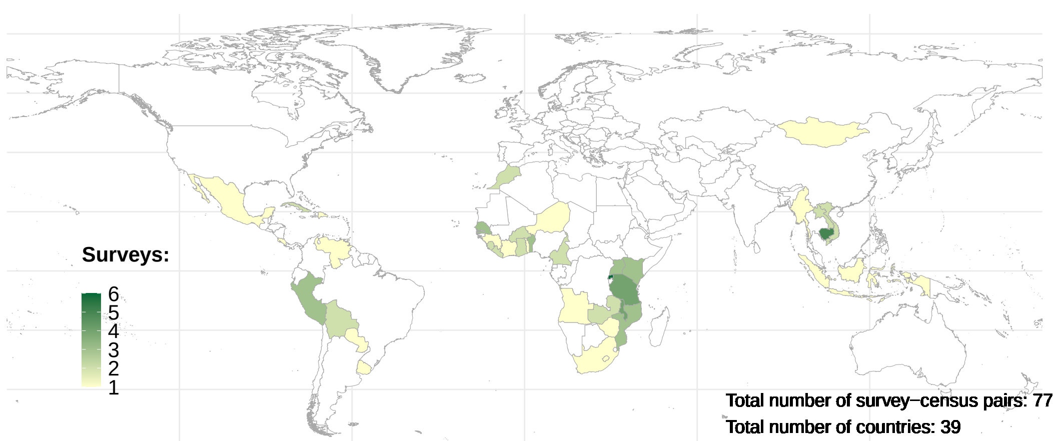


Figure A10: Geographic coverage of DHS/MICS-census pairs

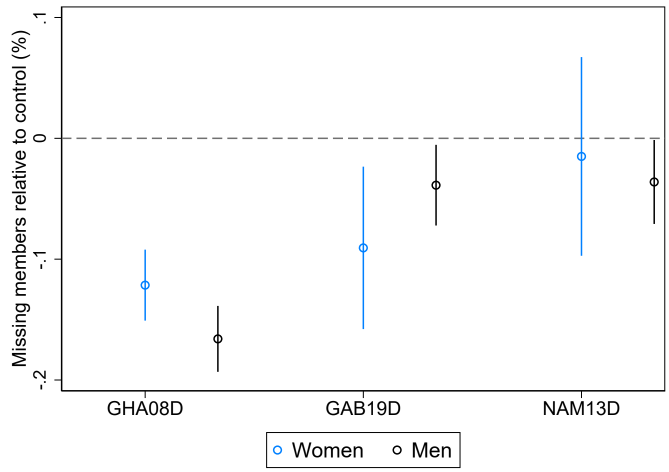
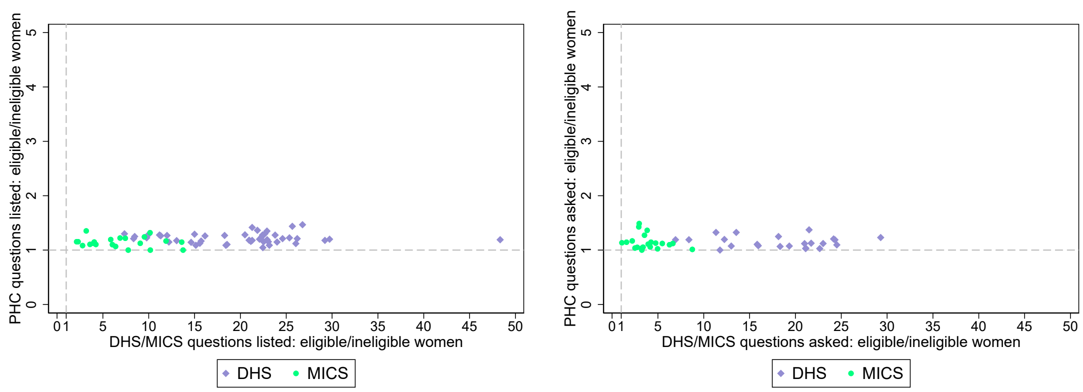


Figure A11: Effect of woman's/man's questionnaire on number of eligible women/men

This figure displays estimates of coefficients from the regression of the eligible number of women (in blue) and men (in black) on the eligibility of their household for the respective individual (woman's or man's) questionnaire. The sample consists of all 3 DHS with a woman's questionnaire that is randomly assigned across households. Circles indicate point estimates and bars indicate 95% confidence intervals. Surveys are sorted along the x-axis in ascending order of the point estimate on the number of eligible women. All surveys are labelled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively. All estimates are reported in Table A5, column (3).



(a) Total number of questions listed

(b) Mean number of questions asked

Figure A12: Question load of eligible women relative to ineligible women

This figure plots the question load of eligible women relative to ineligible women in the DHS/MICS against the same ratio in the matched population and housing censuses (PHC). In Panel (a), question load is measured by the total number of questions listed in the in the roster and the woman's questionnaire. In Panel (b), it is measured by the mean number of question answered about women of eligible and ineligible age. Panel (a) includes data on all 21 MICS-census pairs and all 46 DHS-census pairs. Panel (b) excludes 23 DHS-census pairs. See Appendix A.1.3 for more information.

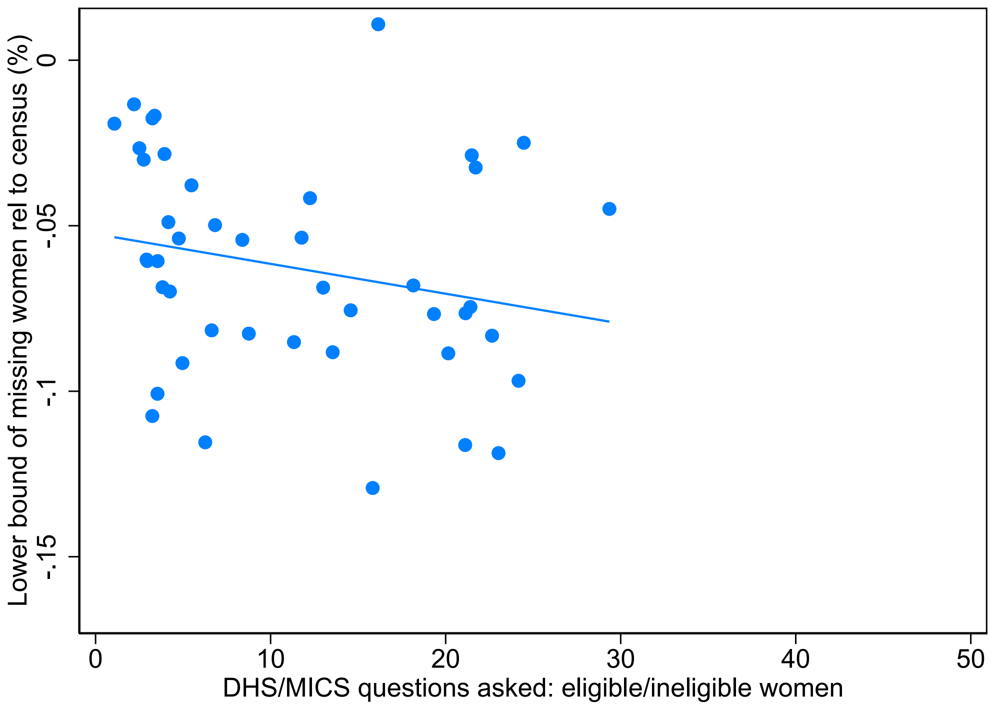


Figure A13: Effect of woman's questionnaire vs. relative question load across surveys

This figure plots estimates of the lower bound of the effect of the woman's questionnaire relative to the control mean against the relative question load faced by eligible women relative to ineligible women across surveys. The sample consists of all 44 DHS and MICS for which the corresponding question counts are available. See appendix A.1.3 for details on the measurement of the length of questionnaires. The solid line presents a linear fit.

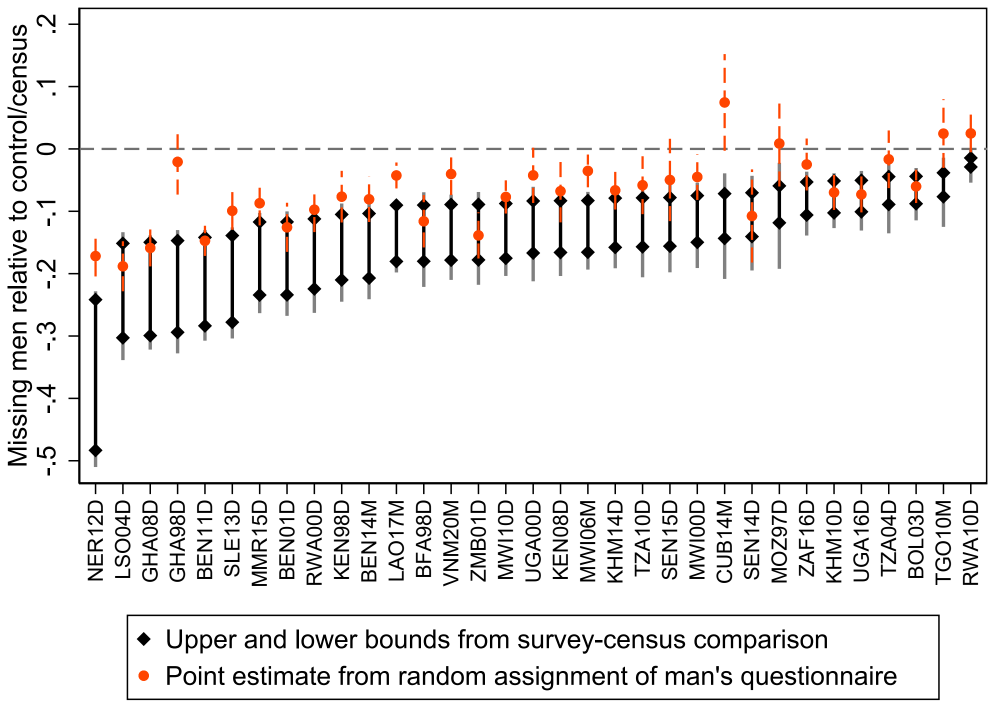


Figure A14: Missing men in DHS/MICS households with a man's questionnaire

This figure displays estimates of the upper and lower bounds of missing men derived using the difference-in-differences approach described in section 4.2.1 as well as estimates of missing men from the comparison of households with and without a man's questionnaire as detailed in section 4.1.1. Black diamonds indicate upper and lower bounds. The area in between bounds is also colored in black. Grey bars indicate 95% confidence intervals of the bounds. Orange circles indicate the point estimates exploiting the random assignment of the man's questionnaire. Dashed orange bars indicate the 95% confidence intervals of these estimates. The sample consists of all 33 surveys for which both estimation approaches are feasible. Surveys are sorted along the x-axis in ascending order of the point estimate of the lower bound. All surveys are labelled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively.

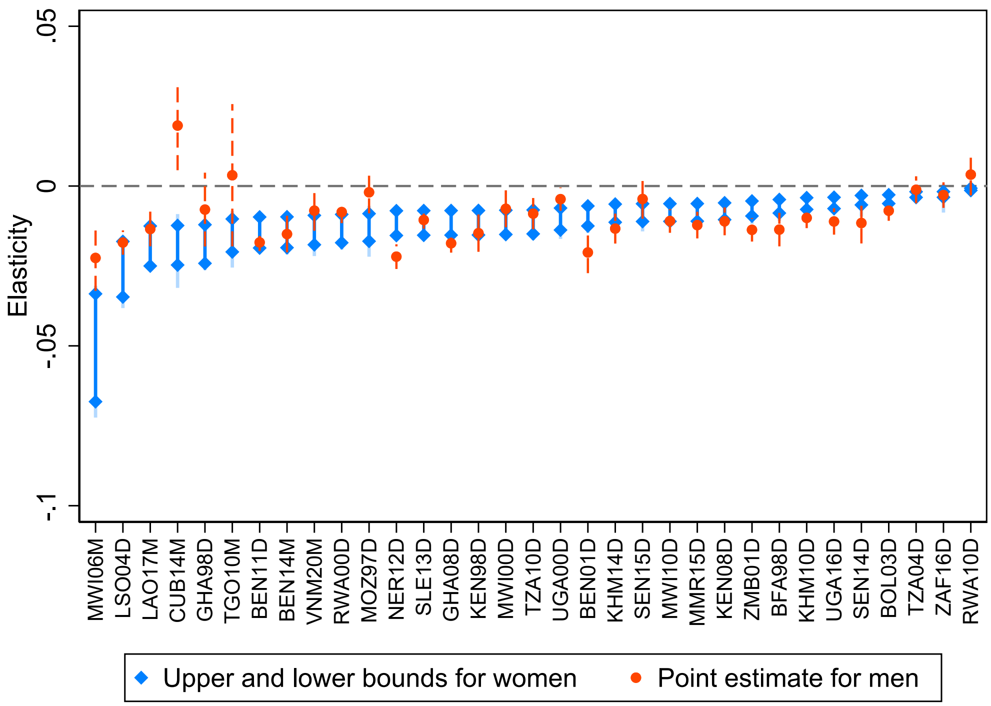


Figure A15: Elasticity of sample size with respect to question load: woman's vs. man's questionnaire

This figure displays estimates of the elasticity of sample size with respect to question load. Upper and lower bounds of the elasticity of sampled eligible women with respect to the question load of women are indicated by blue diamonds. The area between the bounds is also colored in blue. Grey shaded bars indicate the 95% confidence intervals of the bounds. Point estimates of the elasticity of sampled eligible men with respect to the question load of men are indicated by orange circles. Dashed orange bars indicate the 95% confidence intervals of these estimates. The sample consists of all 33 surveys for which the estimation of both elasticities is feasible. Surveys are sorted along the x-axis in ascending order of the point estimate of the lower bound. All surveys are labelled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively.

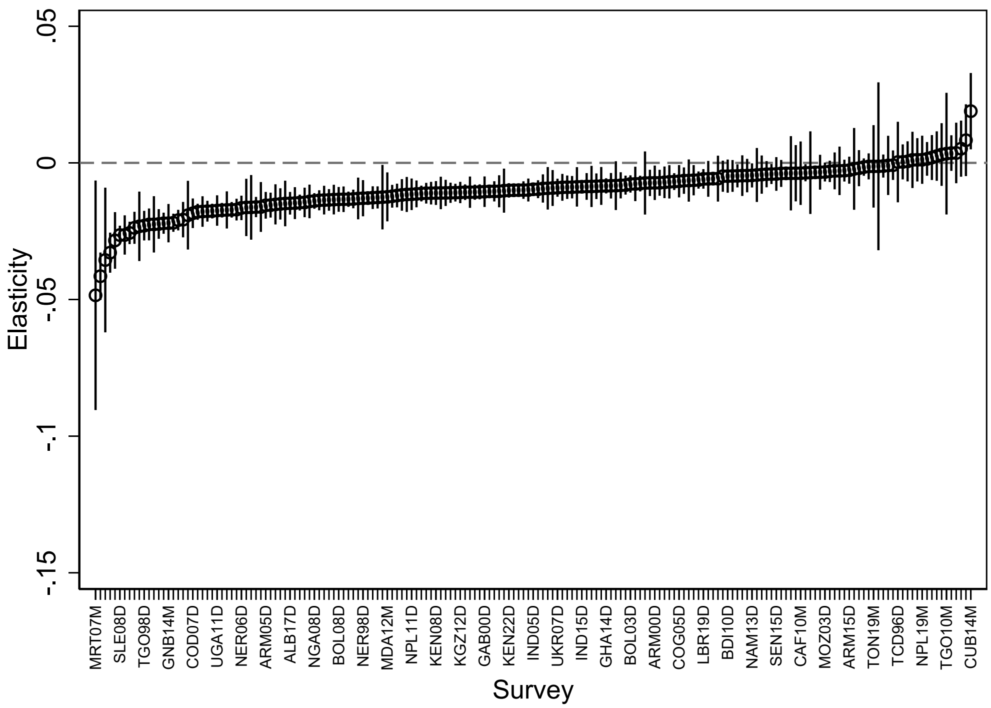


Figure A16: Elasticity of sampled men with respect to question load

This figure displays estimates of the elasticity of sampled men with respect to question load. Point estimates are indicated by black circles. Black bars indicate 95% confidence intervals. The sample consists of all 181 surveys with a man's questionnaire that is randomly assigned across households. Surveys are sorted along the x-axis in ascending order of the point estimate. Every 5th survey is labelled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively.

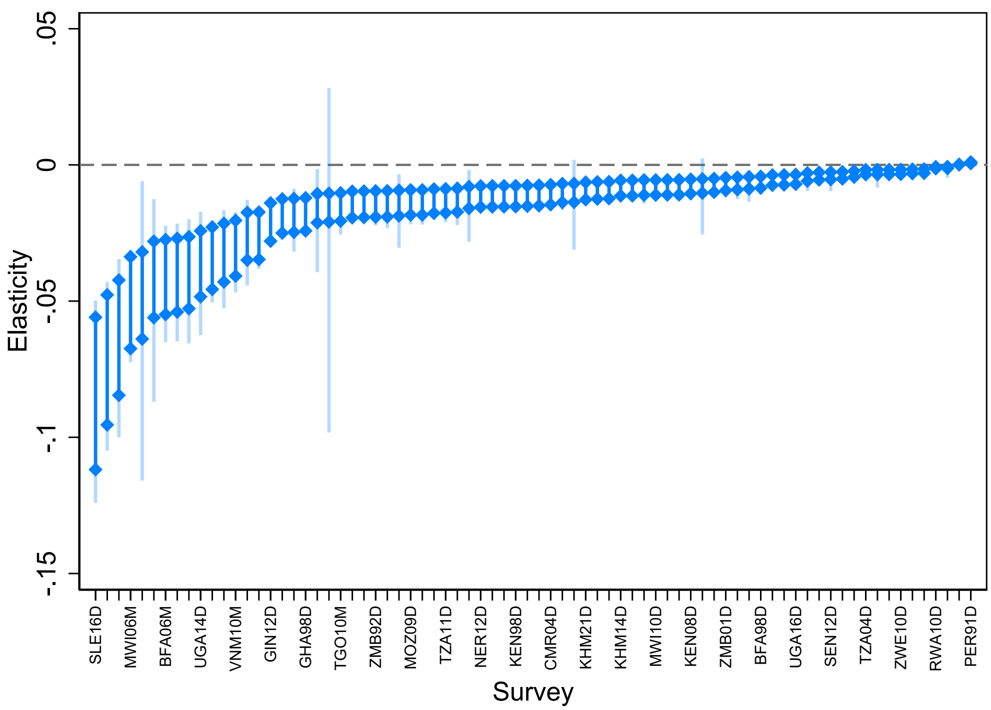


Figure A17: Elasticity of sampled women with respect to question load

This figure displays estimates of the elasticity of sampled women with respect to question load. Upper and lower bounds of the elasticity are indicated by blue diamonds. The area between the bounds is also colored in blue. Grey shaded bars indicate the 95% confidence intervals of the bounds. The sample consists of all 76 survey-census pairs. Surveys are sorted along the x-axis in ascending order of the point estimate. Every 3rd survey is labelled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively.

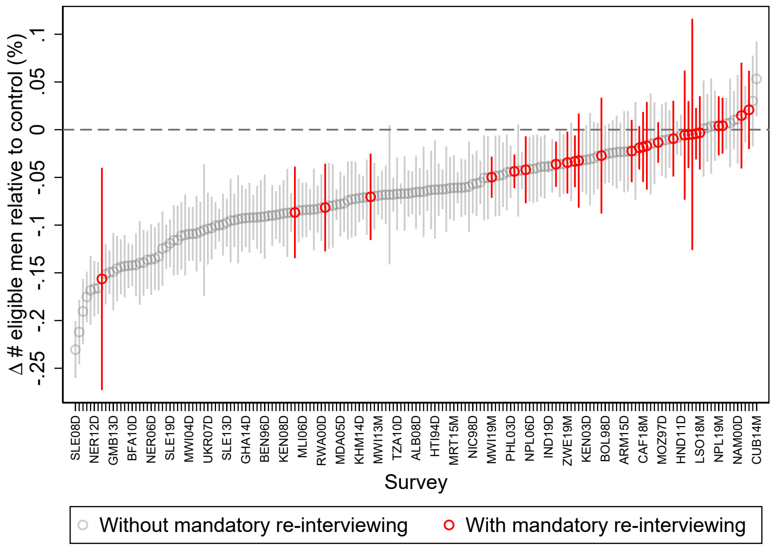
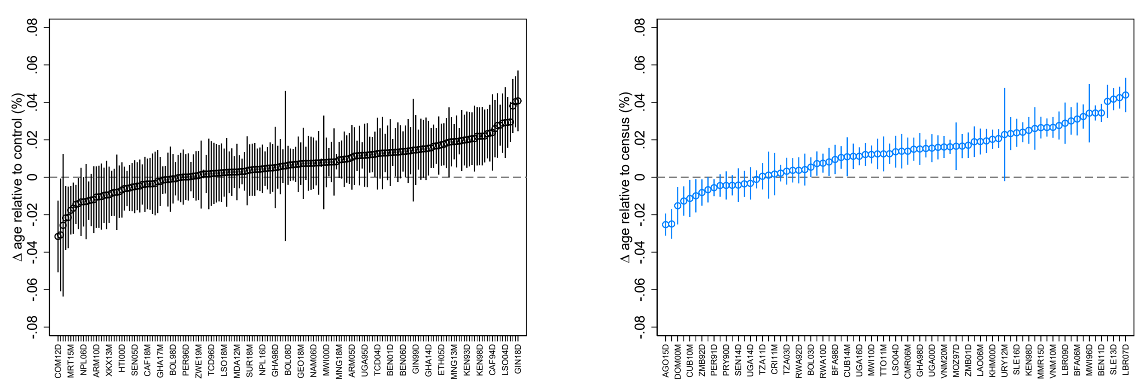


Figure A18: Missing men in surveys with and without mandatory re-interviewing

This figure displays estimates of β from equation (3) relative to the control mean where the outcome variable is the number of eligible men in the household. The sample consists of all 181 DHS and MICS with a man's questionnaire that is randomly assigned across households. Circles indicate point estimates and bars indicate 95% confidence intervals. Estimates from surveys that feature mandatory re-interviewing are shown in red. Surveys are sorted along the x-axis in ascending order of the point estimate. Every 5th survey is labelled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively. All estimates are reported in Table A4, column (3).



(a) Men (DHS/MICS RCT)

(b) Women (census comparison)

Figure A19: Selection on age

This figure displays estimates of the effect of household assignment to the man's (left) and woman's questionnaire (right) on the age of eligible men and women relative to the relevant comparison group, i.e., control households for men and census households for women. Standard errors are clustered at the household-level. Circles indicate point estimates and bars indicate 95% confidence intervals. Surveys are sorted along the x-axis in ascending order of the respective point estimate. In panel (a) every fifth survey is labeled, in panel (b) every 2nd survey is labeled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively. All estimates are reported in Tables A8 and A9, column (2).

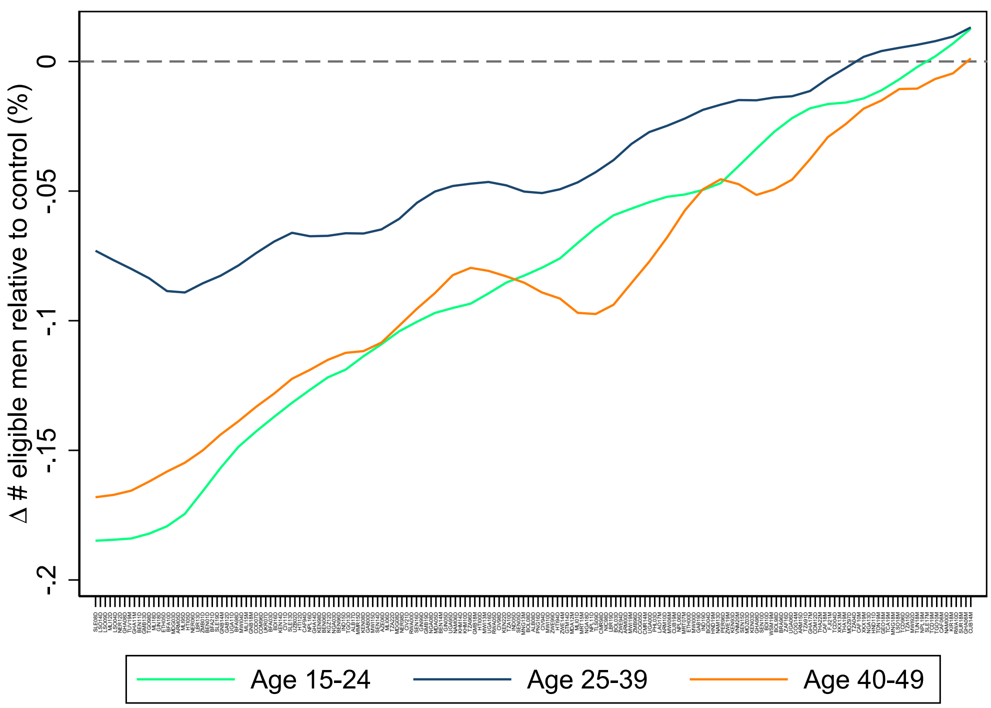
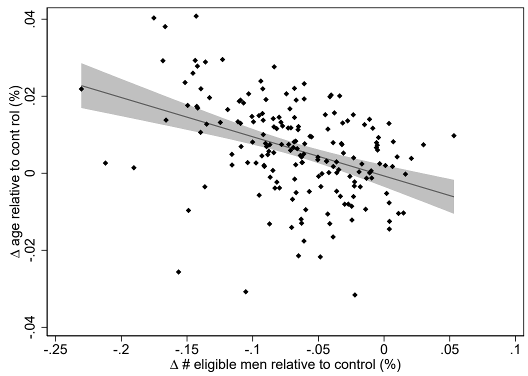
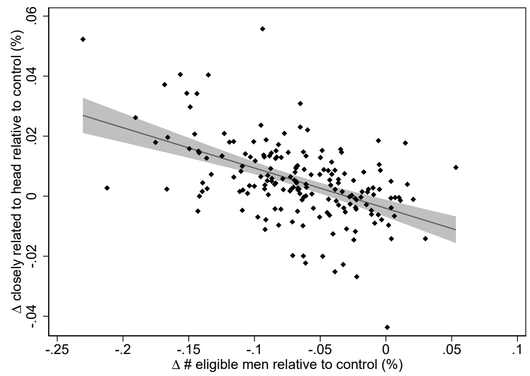


Figure A20: Effect of man's questionnaire on number of eligible men by age group

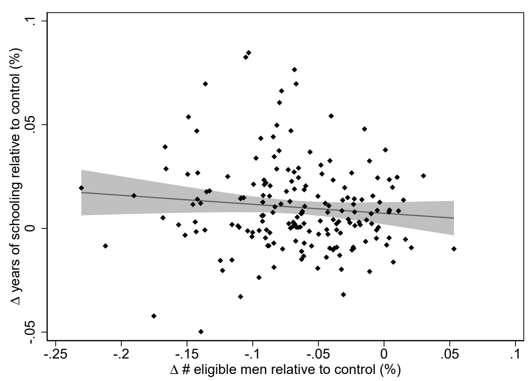
This figure shows the smoothed values from a kernel-weighted local polynomial regression of survey-level regression coefficients of eligible men in the household on the eligibility for the man's questionnaire by age group. The three considered age groups are (i) the ten-year band just above the lower eligibility threshold (typically 15-24), (ii) the 10-year band just below the upper eligibility threshold (typically 40-49) and (iii) the remaining eligible ages in between (typically 25-39).



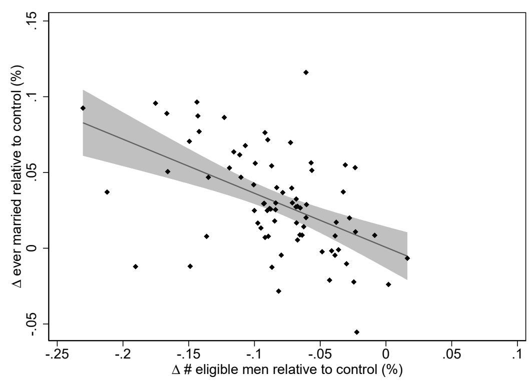
(a) Age



(b) Relationship to household head



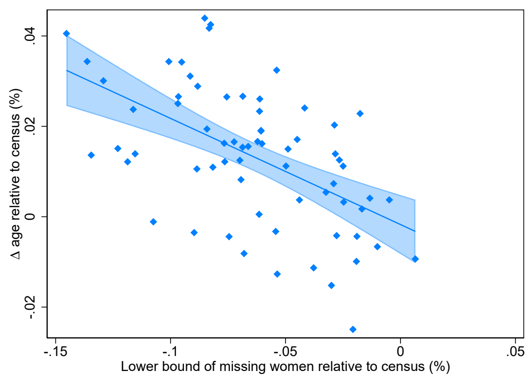
(c) Education



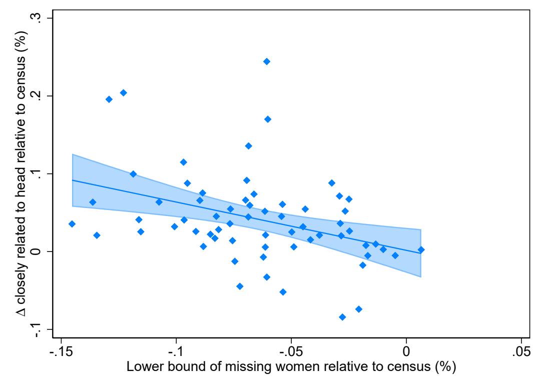
(d) Marital status

Figure A21: Selection on observables vs. missing men

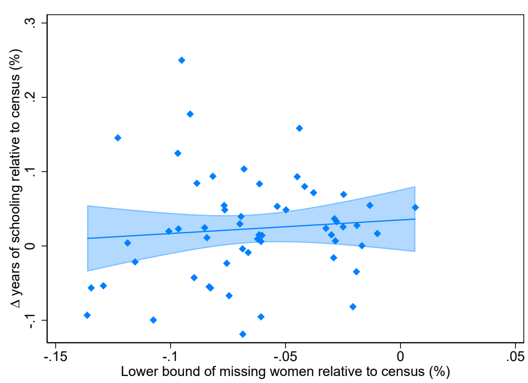
This figure plots coefficients from regressions of individual characteristics of eligible men – age in panel (a), close relationship to household head in panel (b), years of schooling in panel (c), ever married in panel (d) – on household assignment to the man's questionnaire on the y-axis. Coefficients from regressions of the number of eligible men on household assignment to the man's questionnaire are plotted on the x-axis (all panels). Each black dot represents a survey. The grey line represents a linear fit, with the 95% confidence interval indicated by the shaded area.



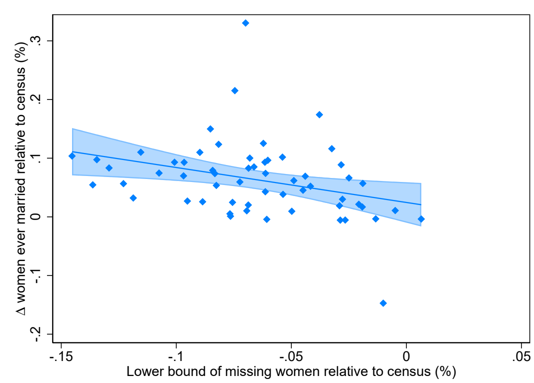
(a) Age



(b) Relationship to household head



(c) Education



(d) Marital status

Figure A22: Selection on observables vs. missing women

This figure plots differences in average individual characteristics of women of eligible age – age in panel (a), close relationship to household head in panel (b), years of schooling in panel (c), ever married in panel (d) – between surveys and matched censuses on the y-axis. The lower bound of missing women is plotted on the x-axis (all panels). Each blue dot represents a survey. The blue line represents a linear fit, with the 95% confidence interval indicated by the shaded area.

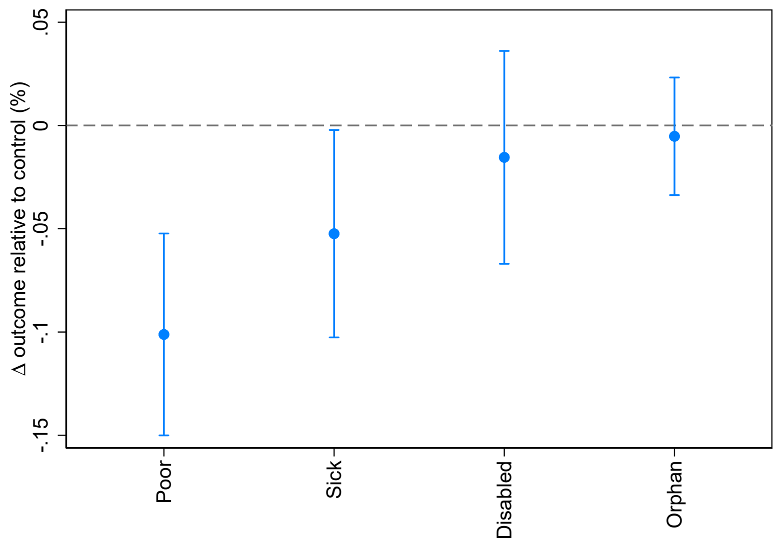


Figure A23: Effect of man's questionnaire on other individual characteristics (DHS)

This figure plots coefficients from regressions of individual characteristics of eligible men on household assignment to the man's questionnaire on the y-axis. Outcomes are indicated on the x-axis. Each regression pools the data from all DHS that include the relevant outcome information. Whiskers indicate the 95% confidence interval. Details on the definitions of the outcomes and the underlying samples are available in appendix A.1.4.

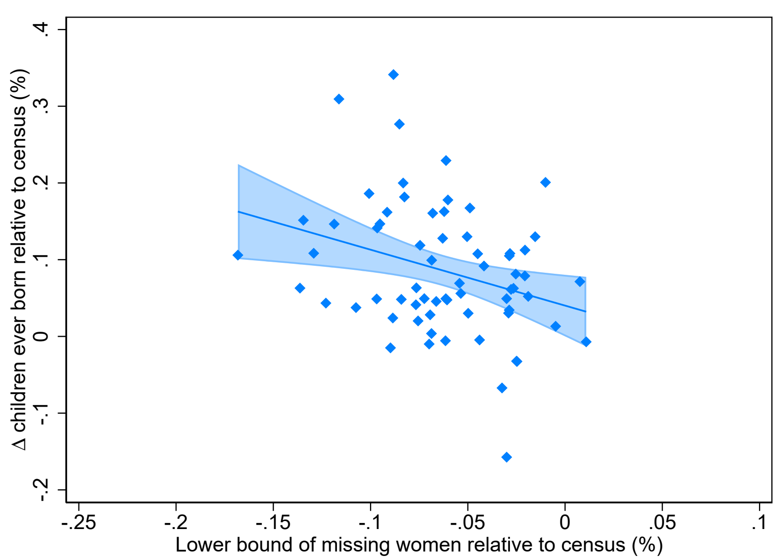
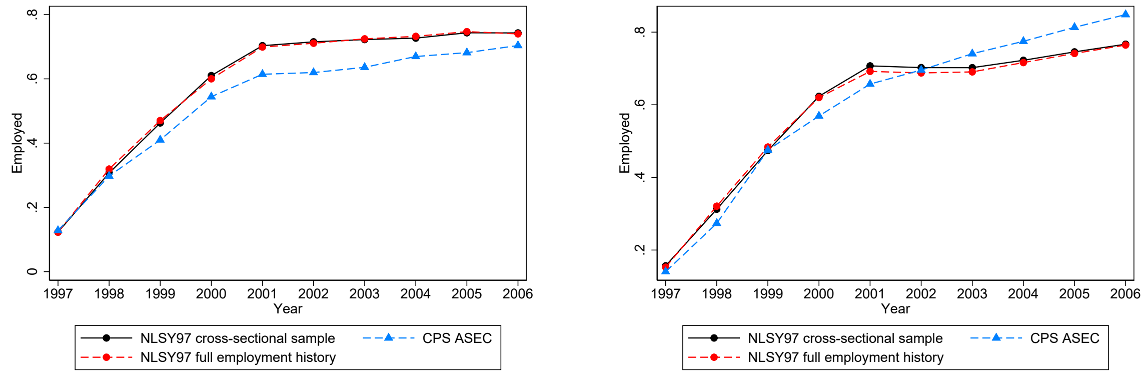


Figure A24: Bias in aggregate fertility vs. missing women

This figure plots differences in the average number of children ever born to women aged 15 to 49 between DHS/MICS and contemporaneous population censuses Each blue dot represents a survey. The blue line represents a linear fit, with the 95% confidence interval indicated by the shaded area.



(a) Female (b) Male

Figure A25: Employment in the NLSY97 vs. CPS ASEC by gender

This figure compares female and male employment rates for the cohorts born in 1980 and 1981 in the NLSY97 and the CPS ASEC. Individuals are considered employment if they have worked at least 520 hours in the past calendar year.

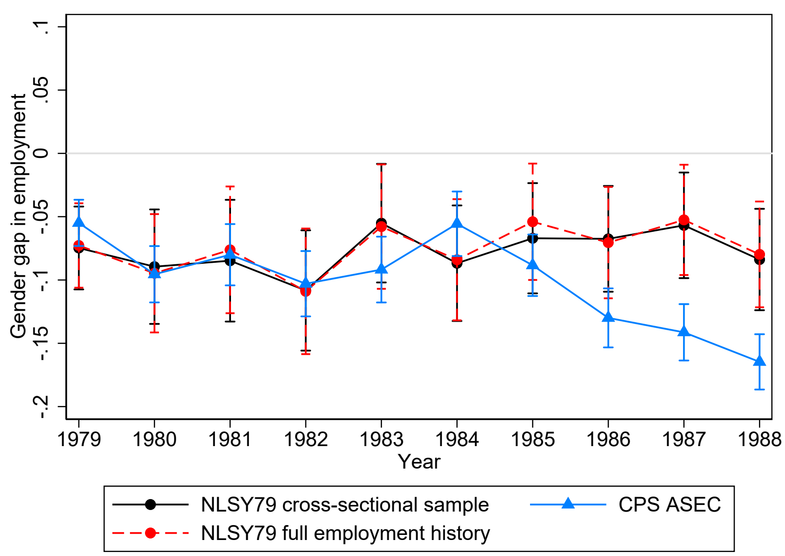


Figure A26: The gender gap in US youth employment: NLSY79 vs. CPS-ASEC

This figure displays estimates of the gender gap in employment for cohorts born in 1962 and 1963 in the NLSY79 and the CPS ASEC over time. Individuals are considered employed if they have worked at least 520 hours in the past calendar year. The NLSY79 full employment history sample is restricted to NLSY79 participants for which a complete weekly employment history is available from 1979 to 1988. Markers indicate point estimates and bars indicate 95% confidence intervals.

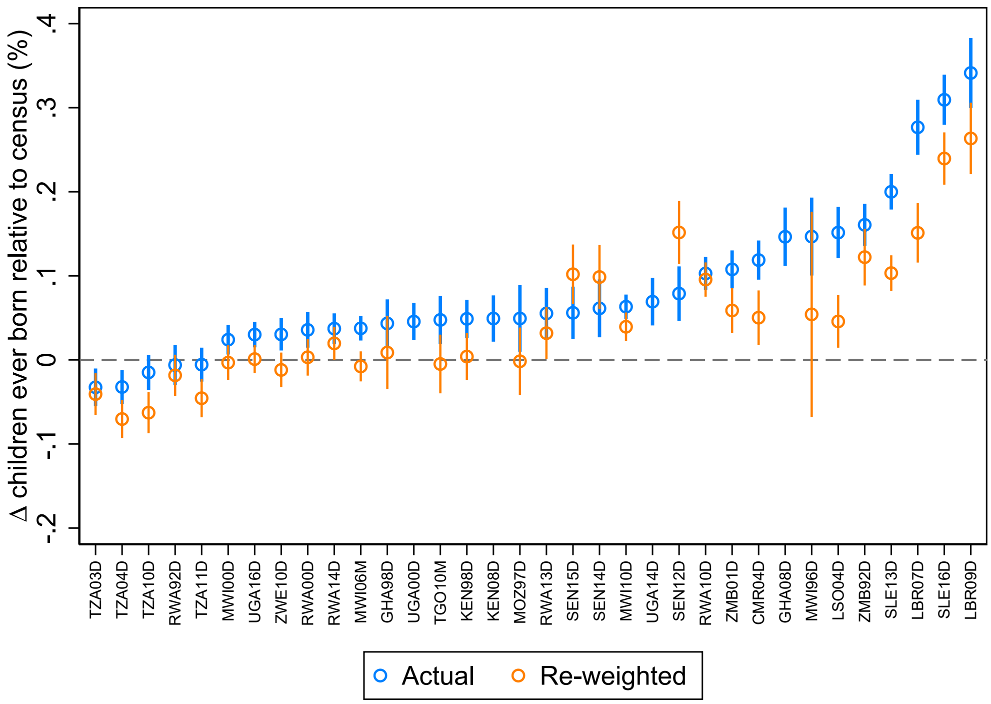
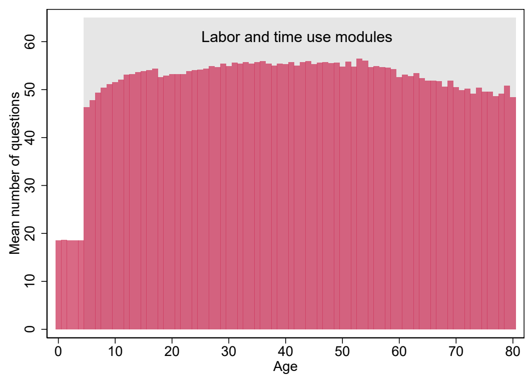


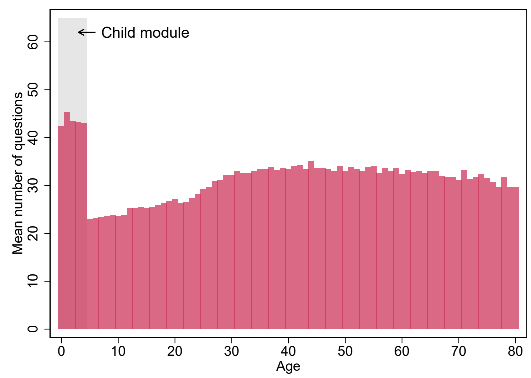
Figure A27: Fertility of women (census comparison) before and after re-weighting

This figure reproduces the effect of household assignment to the woman's questionnaire on the measures of fertility of eligible women relative to the control group of census households for women, as shown in Figure 9a. It adds estimates of fertility using re-weighted DHS/MICS data in orange. The re-weighting procedure applied is 'raking', also called 'matrix scaling' or 'RAS algorithm', a process of iterative proportional fitting whereby data sample weights are adjusted to match desired marginal totals. Raking is performed on all variables present in both survey roster and census data: bins of age, bins of education, relationship to household head and marriage, using the contemporaneous census to provide population marginal distributions to target. In four surveys, empty cells of roster variable combinations prevent raking despite all variables existing in both survey and census. See notes in Figure 9a for further details.

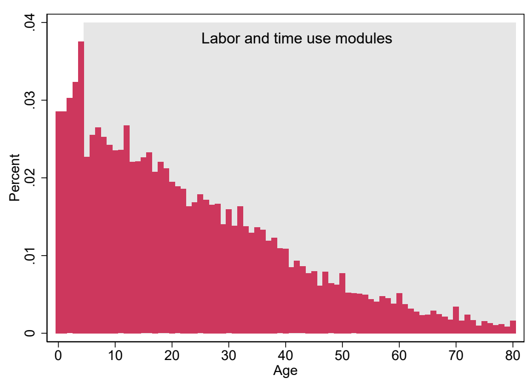
Figure A28: Question load and age distribution in living standards surveys



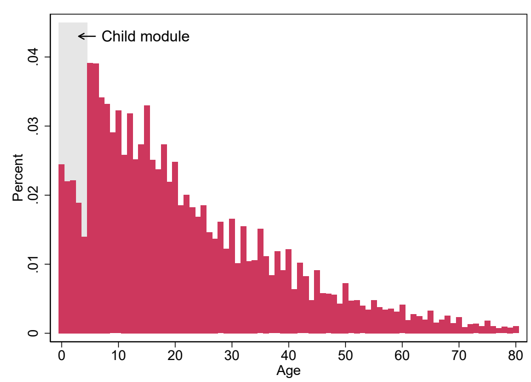
(a) Tanzania HBS 2011: Question load



(b) Zambia LCMS 2015: Question load



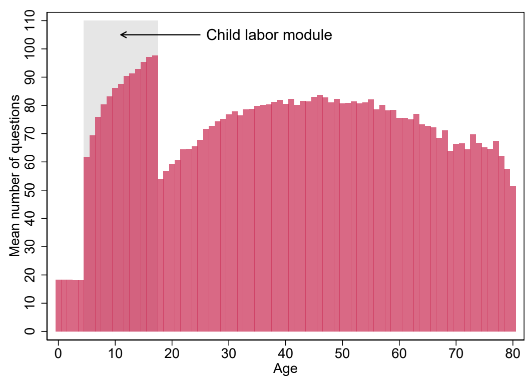
(c) Tanzania HBS 2011: Age distribution



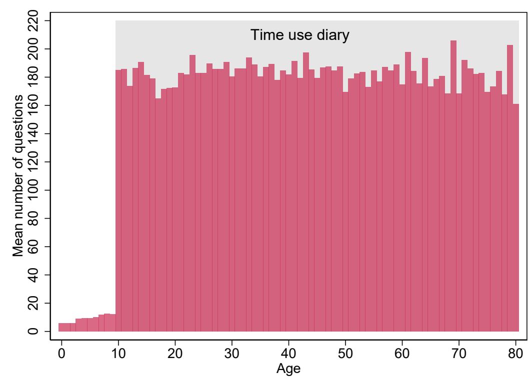
(d) Zambia LCMS 2015: Age distribution

This figure shows the distribution of the mean number of questions asked about household members by age in the 2011 Tanzanian Household Budget Survey (HBS) and the 2015 Zambian Living Conditions Monitoring Survey (LCMS) in the top panels (a) and (b). The bottom panels (c) and (d) show the age distributions in the two surveys. Shaded areas indicate survey modules that are only applied to specific age groups.

Figure A29: Question load and age distribution in labour and time use surveys

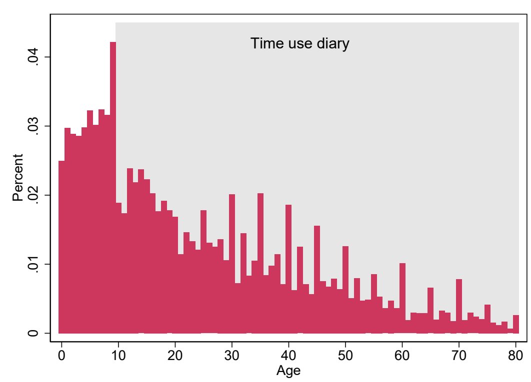


(a) Tanzania ILFS 2014: Question load



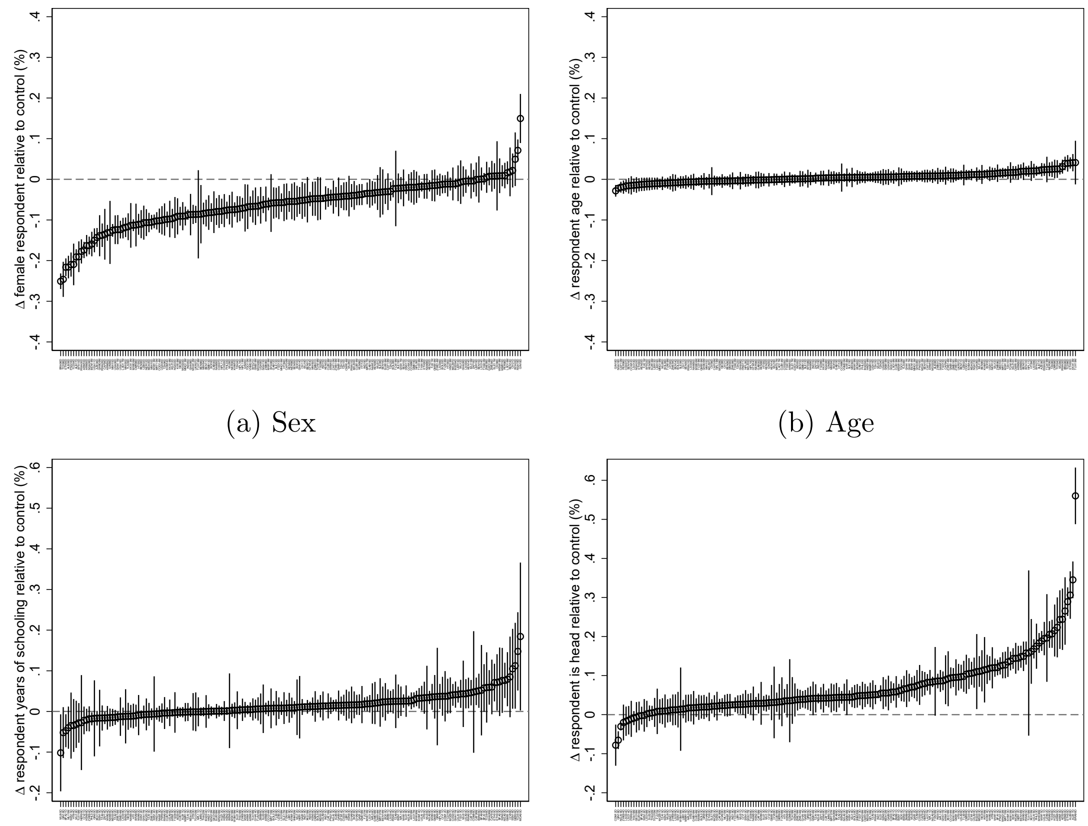
(b) GHA TUS 2009: Question load

(c) Tanzania ILFS 2014: Age distribution



(d) Ghana TUS 2009: Age distribution

This figure shows the distribution of the mean number of questions asked about household members by age in the 2014 Tanzanian Integrated Labour Force Survey (ILFS) and the 2009 Ghana Time Use Survey (TUS) in the top panels (a) and (b). The bottom panels (c) and (d) show the age distributions in the two surveys. Shaded areas indicate survey modules that are only applied to specific age groups.



(a) Sex

(b) Age

(c) Education

(d) Household head

Figure A30: Effect of man's questionnaire on respondent characteristics in the DHS

This figure displays estimates of the effect of household assignment to the man's questionnaire on the characteristics of the respondents to the household roster relative to the control group. Standard errors are clustered at the household-level. Circles indicate point estimates and bars indicate 95% confidence intervals. Surveys are sorted along the x-axis in ascending order of the respective point estimate. All surveys are labeled. Survey labels are composed of three-letter country codes, followed by the year of the survey and a single letter D or M standing for DHS or MICS, respectively.

A.6 Appendix Tables

Table A1: Surveys with randomly assigned man's questionnaire excluded from analysis

Reason for exclusion	**Excluded surveys**	**Total**
Additional survey features administered in control households (without man's questionnaire) that were not implemented in treatment households (with man's questionnaire)	AGO DHS 2015; BEN DHS 2017; CIV MICS 2016; CMR DHS 2004, 2011, 2018; COD DHS 2013; COD MICS 2017; COG DHS 2011; COM MICS 2022; DOM DHS 2002; GIN DHS 2012; JOR DHS 2017; KAZ DHS 1999; KHM DHS 2005, 2021; MDG DHS 2021; MDG MICS 2018; MOZ DHS 2011; MRT DHS 2019; NPL DHS 2022; RWA DHS 2014, 2019; SEN DHS 2018, 2018, 2019; TCD DHS 2014; TLS DHS 2016	28
Eligibility for man's questionnaire conditional on marital status	AFG DHS 2015; BGD DHS 1996, 1999, 2007, 2011; IDN DHS 2002, 2007, 2012, 2017; MDV DHS 2009; NPL DHS 2001; PAK DHS 2012, 2017	13
Randomisation of man's questionnaire stratified by presence of children at household listing stage, but stratification variable not available in microdata	BLR MICS 2012, 2019; GUY MICS 2014; MNE MICS 2013, 2018; UKR MICS 2012	6
No upper age limit for eligibility for man's questionnaire	BFA DHS 1993, MAR DHS 1992, SEN DHS 1992	3
Individual identifiers do not match across microdata source files	STP MICS 2019; SWZ MICS 2014; TGO MICS 2017	3
Random assignment of man's questionnaire across clusters rather than across households within clusters	GHA DHS 1993	1
Contradicting information about assignment of man's questionnaire in survey report and microdata	KAZ MICS 2010	1

Table A2: DHS and MICS with randomly assigned man's questionnaire

Country code	Country name	DHS	MICS
ALB	Albania	2008, 2017	NA
ARM	Armenia	2000, 2005, 2010, 2015	NA
AZE	Azerbaijan	2006	NA
BDI	Burundi	2010, 2016	NA
BEN	Benin	1996, 2001, 2006, 2011	2014
BFA	Burkina Faso	1998, 2003, 2010, 2021	NA
BGD	Bangladesh	2004	NA
BOL	Bolivia	1998, 2003, 2008	NA
BRA	Brazil	1996	NA
CAF	Central African Republic	1994	2006, 2010, 2018
CIV	Côte d'Ivoire	1994, 1998, 2011, 2021	NA
CMR	Cameroon	1998	2014
COD	Congo - Kinshasa	2007	NA
COG	Congo - Brazzaville	2005	2014
COM	Comoros	1996, 2012	NA
CUB	Cuba	NA	2014, 2019
ETH	Ethiopia	2000, 2005	NA
FJI	Fiji	NA	2021
GAB	Gabon	2000, 2012, 2019	NA
GEO	Georgia	NA	2018
GHA	Ghana	1998, 2008, 2014	2006, 2011, 2017
GIN	Guinea	1999, 2005, 2018	NA
GMB	Gambia	2013, 2019	2018
GNB	Guinea-Bissau	NA	2014, 2018
GTM	Guatemala	2014	NA
HND	Honduras	2011	2019
HTI	Haiti	1994, 2000, 2005, 2012	NA
IND	India	2005, 2015, 2019	NA
KEN	Kenya	1993, 1998, 2003, 2008, 2014, 2022	NA
KGZ	Kyrgyzstan	2012	NA
KHM	Cambodia	2010, 2014	NA
KIR	Kiribati	NA	2018
LAO	Laos	NA	2017
LBR	Liberia	2013, 2019	NA
LSO	Lesotho	2004, 2009, 2014	2018
MDA	Moldova	2005	2012
MDG	Madagascar	2003, 2008	NA
MLI	Mali	1995, 2001, 2006, 2012, 2018	2015
MMR	Myanmar (Burma)	2015	NA
MNG	Mongolia	NA	2013, 2018
MOZ	Mozambique	1997, 2003	NA
MRT	Mauritania	NA	2007, 2015
MWI	Malawi	1992, 2000, 2004, 2010, 2015	2006, 2013, 2019
NAM	Namibia	2000, 2006, 2013	NA
NER	Niger	1998, 2006, 2012	NA
NGA	Nigeria	2003, 2008, 2013, 2018	NA
NIC	Nicaragua	1998	NA
NPL	Nepal	2006, 2011, 2016	2019
PER	Peru	1996	NA
PHL	Philippines	2003	NA
PNG	Papua New Guinea	2016	NA
RWA	Rwanda	2000, 2005, 2010	NA
SEN	Senegal	2005, 2010, 2014, 2015, 2016	NA
SLE	Sierra Leone	2008, 2013, 2019	2017
SUR	Suriname	NA	2018
TCA	Turks and Caicos Islands	NA	2019
TCD	Chad	1996, 2004	2019
TGO	Togo	1998, 2013	2010
THA	Thailand	NA	2019, 2022
TLS	Timor-Leste	2009	NA
TON	Tonga	NA	2019
TUN	Tunisia	NA	2018
TUV	Tuvalu	NA	2019
TZA	Tanzania	1991, 1996, 2004, 2010, 2015, 2022	NA
UGA	Uganda	1995, 2000, 2006, 2011, 2016	NA
UKR	Ukraine	2007	NA
UZB	Uzbekistan	2002	NA
VNM	Vietnam	NA	2020
WSM	Samoa	NA	2019
XKX	Republic of Kosovo	NA	2013, 2019
ZAF	South Africa	2016	NA
ZMB	Zambia	1996, 2001	NA
ZWE	Zimbabwe	1994, 1999	2014, 2019

Table A3: MICS/DHS-Population Census pairs

Country	Survey	Survey Year	PHC Year	Source	Statistical Office
BEN	DHS	2001	2002	IPUMS	National Institute of Statistics and Economic Analysis
BEN	DHS	2011	2013	IPUMS	National Institute of Statistics and Economic Analysis
BEN	MICS	2014	2013	IPUMS	National Institute of Statistics and Economic Analysis
BFA	MICS	2006	2006	IPUMS	National Institute of Statistics and Demography
BOL	DHS	1994	1992	IPUMS	National Institute of Statistics
BOL	DHS	2003	2001	IPUMS	National Institute of Statistics
CMR	DHS	2004	2005	IPUMS	Central Bureau of Census and Population Studies
CMR	MICS	2006	2005	IPUMS	Central Bureau of Census and Population Studies
CRI	MICS	2011	2011	IPUMS	National Institute of Statistics and Censuses
CUB	MICS	2010	2012	IPUMS	Office of National Statistics
CUB	MICS	2014	2012	IPUMS	Office of National Statistics
DOM	MICS	2000	2002	IPUMS	National Statistics Office
GHA	DHS	1998	2000	IPUMS	Ghana Statistical Services
GHA	DHS	2008	2010	IPUMS	Ghana Statistical Services
IDN	MICS	2000	2000	IPUMS	Statistics Indonesia
KEN	DHS	1989	1989	IPUMS	National Bureau of Statistics
KEN	DHS	1998	1999	IPUMS	National Bureau of Statistics
KEN	DHS	2008	2009	IPUMS	National Bureau of Statistics
KHM	DHS	2000	1998	IPUMS	National Institute of Statistics
KHM	DHS	2010	2008	IPUMS	National Institute of Statistics
KHM	DHS	2014	2013	IPUMS	National Institute of Statistics
KHM	DHS	2021	2019	IPUMS	National Institute of Statistics
LAO	MICS	2006	2005	IPUMS	Statistics Bureau
LAO	MICS	2017	2015	IPUMS	Statistics Bureau
LBR	DHS	2007	2008	IPUMS	Institute of Statistics and Geo-Information Systems
LBR	DHS	2009	2008	IPUMS	Institute of Statistics and Geo-Information Systems
LSO	DHS	2004	2006	IPUMS	Bureau of Statistics
MEX	MICS	2015	2015	IPUMS	National Institute of Statistics, Geography, and Informatics
MMR	DHS	2015	2014	IPUMS	Central Statistical Organization
MNG	MICS	2010	2010	NSO	National Statistical Office
MOZ	DHS	1997	1997	IPUMS	National Institute of Statistics
MOZ	MICS	2008	2007	IPUMS	National Institute of Statistics
MOZ	DHS	2009	2007	IPUMS	National Institute of Statistics
MWI	DHS	1996	1998	IPUMS	National Statistical Office
MWI	DHS	2000	1998	IPUMS	National Statistical Office
MWI	MICS	2006	2008	IPUMS	National Statistical Office
MWI	DHS	2010	2008	IPUMS	National Statistical Office
NER	DHS	2012	2012	NSO	National Institute of Statistics
PER	DHS	1991	1993	IPUMS	National Institute of Statistics and Informatics
PER	DHS	2007	2007	IPUMS	National Institute of Statistics and Informatics
PER	DHS	2009	2007	IPUMS	National Institute of Statistics and Informatics
PRY	DHS	1990	1992	IPUMS	General Directorate of Statistics, Surveys, and Censuses
RWA	DHS	1992	1991	IPUMS	National Institute of Statistics
RWA	DHS	2000	2002	IPUMS	National Institute of Statistics
RWA	MICS	2000	2002	IPUMS	National Institute of Statistics
SEN	DHS	2012	2013	IPUMS	National Agency of Statistics and Demography
SEN	DHS	2014	2013	IPUMS	National Agency of Statistics and Demography
SEN	DHS	2015	2013	IPUMS	National Agency of Statistics and Demography
SLE	DHS	2013	2015	IPUMS	Statistics Sierra Leone
SLE	DHS	2016	2015	IPUMS	Statistics Sierra Leone
TGO	MICS	2010	2010	IPUMS	National Institute of Statistics (INSEED)
TTO	MICS	2011	2011	IPUMS	Central Statistical Office
TZA	DHS	2003	2002	IPUMS	National Bureau of Statistics
TZA	DHS	2004	2002	IPUMS	National Bureau of Statistics
TZA	DHS	2010	2012	IPUMS	National Bureau of Statistics
TZA	DHS	2011	2012	IPUMS	National Bureau of Statistics
UGA	DHS	2000	2002	IPUMS	Bureau of Statistics
UGA	DHS	2014	2014	IPUMS	Bureau of Statistics
UGA	DHS	2016	2014	IPUMS	Bureau of Statistics
URY	MICS	2012	2011	IPUMS	National Institute of Statistics
VEN	MICS	2000	2001	IPUMS	National Institute of Statistics
VNM	MICS	2010	2009	IPUMS	General Statistics Office
VNM	MICS	2020	2019	IPUMS	General Statistics Office
ZAF	DHS	2016	2016	IPUMS	Statistics South Africa
ZMB	DHS	1992	1990	IPUMS	Central Statistical Office
ZMB	DHS	2001	2000	IPUMS	Central Statistical Office
ZWE	DHS	2010	2012	IPUMS	Central Statistical Office

Table A4: Effect of man's questionnaire on number of men in the household

Survey	Eligible men Absolute	Eligible men Relative	Ineligible men Absolute	Ineligible men Relative	Total men Absolute	N
ALB DHS 2008	-0.055 (0.017)	-0.065 (0.020)	0.009 (0.014)	0.012 (0.018)	-0.046 (0.020)	7,999
ALB DHS 2017	-0.089 (0.013)	-0.089 (0.013)	0.001 (0.009)	0.001 (0.017)	-0.088 (0.014)	15,823
ARM DHS 2000	-0.051 (0.023)	-0.050 (0.022)	0.023 (0.022)	0.031 (0.029)	-0.028 (0.028)	5,980
ARM DHS 2005	-0.120 (0.020)	-0.140 (0.022)	0.014 (0.018)	0.020 (0.027)	-0.106 (0.025)	6,705
ARM DHS 2010	-0.034 (0.020)	-0.043 (0.025)	-0.015 (0.016)	-0.023 (0.026)	-0.049 (0.025)	6,700
ARM DHS 2015	-0.017 (0.016)	-0.023 (0.022)	0.006 (0.014)	0.008 (0.020)	-0.012 (0.021)	7,893
AZE DHS 2006	-0.105 (0.022)	-0.085 (0.017)	0.073 (0.016)	0.160 (0.038)	-0.031 (0.025)	7,171
BDI DHS 2010	-0.030 (0.020)	-0.028 (0.018)	0.003 (0.015)	0.006 (0.032)	-0.028 (0.025)	8,593
BDI DHS 2016	-0.108 (0.014)	-0.101 (0.013)	-0.028 (0.011)	-0.054 (0.021)	-0.137 (0.019)	15,977
BEN DHS 1996	-0.091 (0.023)	-0.091 (0.022)	0.018 (0.035)	0.019 (0.038)	-0.073 (0.044)	4,498
BEN DHS 2001	-0.141 (0.024)	-0.125 (0.020)	0.064 (0.022)	0.116 (0.043)	-0.077 (0.035)	5,768
BEN DHS 2006	-0.101 (0.014)	-0.092 (0.012)	0.040 (0.013)	0.078 (0.025)	-0.061 (0.020)	17,489
BEN DHS 2011	-0.167 (0.014)	-0.150 (0.012)	0.061 (0.013)	0.117 (0.025)	-0.106 (0.019)	17,422
BEN MICS 2014	-0.082 (0.016)	-0.077 (0.015)	-0.010 (0.015)	-0.015 (0.024)	-0.092 (0.023)	14,073
BFA DHS 1998	-0.148 (0.031)	-0.110 (0.022)	0.043 (0.027)	0.056 (0.037)	-0.106 (0.043)	4,812
BFA DHS 2003	-0.144 (0.025)	-0.103 (0.017)	0.032 (0.021)	0.042 (0.028)	-0.112 (0.035)	9,093
BFA DHS 2010	-0.173 (0.015)	-0.142 (0.011)	0.002 (0.014)	0.004 (0.022)	-0.170 (0.021)	14,423
BFA DHS 2021	-0.167 (0.017)	-0.123 (0.012)	0.038 (0.016)	0.054 (0.022)	-0.129 (0.024)	13,251
BGD DHS 2004	-0.048 (0.019)	-0.038 (0.014)	-0.003 (0.016)	-0.005 (0.022)	-0.051 (0.024)	10,500
BOL DHS 1998	-0.028 (0.017)	-0.026 (0.015)	0.031 (0.014)	0.065 (0.029)	0.003 (0.021)	12,106
BOL DHS 2003	-0.060 (0.013)	-0.055 (0.012)	0.024 (0.010)	0.055 (0.023)	-0.036 (0.016)	19,204
BOL DHS 2008	-0.069 (0.012)	-0.066 (0.011)	0.004 (0.010)	0.008 (0.022)	-0.066 (0.015)	19,561
BRA DHS 1996	-0.029 (0.019)	-0.025 (0.016)	0.035 (0.014)	0.074 (0.030)	0.006 (0.022)	13,274
CAF DHS 1994	-0.103 (0.027)	-0.094 (0.024)	0.001 (0.024)	0.001 (0.042)	-0.102 (0.036)	5,551
CAF MICS 2006	0.009 (0.014)	0.010 (0.016)	0.024 (0.013)	0.054 (0.030)	0.033 (0.019)	11,721
CAF MICS 2010	-0.009 (0.015)	-0.010 (0.015)	0.007 (0.013)	0.015 (0.029)	-0.002 (0.020)	11,755
CAF MICS 2018	-0.020 (0.021)	-0.018 (0.019)	0.018 (0.018)	0.030 (0.030)	-0.002 (0.027)	8,133
CIV DHS 1994	-0.098 (0.037)	-0.065 (0.024)	-0.004 (0.030)	-0.006 (0.043)	-0.101 (0.052)	5,935
CIV DHS 1998	-0.103 (0.057)	-0.068 (0.037)	0.013 (0.044)	0.020 (0.067)	-0.089 (0.077)	2,122
CIV DHS 2011	-0.128 (0.023)	-0.099 (0.017)	0.035 (0.016)	0.062 (0.030)	-0.093 (0.028)	9,682
CIV DHS 2021	-0.096 (0.015)	-0.083 (0.013)	0.016 (0.012)	0.034 (0.026)	-0.080 (0.020)	14,766
CMR DHS 1998	-0.077 (0.032)	-0.060 (0.024)	0.040 (0.025)	0.063 (0.042)	-0.037 (0.043)	4,693
CMR MICS 2014	-0.047 (0.018)	-0.047 (0.017)	0.018 (0.016)	0.032 (0.029)	-0.029 (0.024)	10,212
COD DHS 2007	-0.134 (0.021)	-0.107 (0.016)	-0.024 (0.017)	-0.040 (0.028)	-0.158 (0.027)	8,885
COG DHS 2005	-0.060 (0.028)	-0.048 (0.022)	0.076 (0.022)	0.137 (0.042)	0.015 (0.036)	5,879
COG MICS 2014	-0.020 (0.014)	-0.023 (0.016)	0.007 (0.013)	0.014 (0.025)	-0.013 (0.018)	12,811
COM DHS 1996	-0.143 (0.050)	-0.105 (0.035)	0.080 (0.045)	0.101 (0.058)	-0.063 (0.068)	2,252
COM DHS 2012	-0.027 (0.032)	-0.022 (0.026)	0.083 (0.028)	0.121 (0.043)	0.056 (0.042)	4,481
CUB MICS 2014	0.039 (0.014)	0.053 (0.020)	-0.015 (0.012)	-0.030 (0.024)	0.024 (0.017)	9,494
CUB MICS 2019	-0.028 (0.012)	-0.042 (0.018)	0.008 (0.011)	0.015 (0.020)	-0.020 (0.015)	11,966
ETH DHS 2000	-0.044 (0.020)	-0.040 (0.018)	0.020 (0.017)	0.033 (0.029)	-0.025 (0.026)	14,071
ETH DHS 2005	-0.169 (0.015)	-0.143 (0.012)	0.038 (0.013)	0.065 (0.024)	-0.131 (0.020)	13,705
FJI MICS 2021	-0.017 (0.024)	-0.017 (0.023)	0.012 (0.019)	0.018 (0.028)	-0.005 (0.030)	5,467
GAB DHS 2000	-0.106 (0.029)	-0.087 (0.024)	0.037 (0.024)	0.056 (0.037)	-0.069 (0.039)	6,203
GAB DHS 2012	-0.118 (0.021)	-0.116 (0.019)	0.037 (0.016)	0.076 (0.034)	-0.082 (0.026)	9,750
GAB DHS 2019	-0.039 (0.017)	-0.039 (0.017)	0.032 (0.012)	0.089 (0.037)	-0.007 (0.022)	11,781
GEO MICS 2018	-0.004 (0.013)	-0.005 (0.018)	0.010 (0.011)	0.015 (0.017)	0.006 (0.016)	12,270
GHA DHS 1998	-0.025 (0.020)	-0.031 (0.025)	-0.004 (0.019)	-0.010 (0.043)	-0.030 (0.028)	6,003
GHA MICS 2006	0.029 (0.023)	0.030 (0.024)	-0.019 (0.021)	-0.033 (0.036)	0.010 (0.032)	5,932
GHA DHS 2008	-0.163 (0.015)	-0.166 (0.014)	0.021 (0.013)	0.048 (0.030)	-0.142 (0.020)	11,778
GHA MICS 2011	-0.138 (0.016)	-0.151 (0.016)	0.000 (0.016)	0.001 (0.024)	-0.138 (0.022)	11,924
GHA DHS 2014	-0.079 (0.014)	-0.093 (0.016)	0.018 (0.012)	0.042 (0.029)	-0.061 (0.019)	11,834
GHA MICS 2017	-0.020 (0.015)	-0.022 (0.017)	0.018 (0.014)	0.029 (0.022)	-0.001 (0.020)	12,886
GIN DHS 1999	-0.114 (0.034)	-0.080 (0.023)	0.040 (0.028)	0.053 (0.038)	-0.073 (0.048)	5,089
GIN DHS 2005	-0.086 (0.025)	-0.074 (0.021)	0.029 (0.023)	0.039 (0.032)	-0.058 (0.036)	6,280
GIN DHS 2018	-0.179 (0.023)	-0.143 (0.017)	-0.003 (0.020)	-0.005 (0.027)	-0.183 (0.031)	7,912
GMB DHS 2013	-0.274 (0.041)	-0.149 (0.021)	0.012 (0.029)	0.013 (0.031)	-0.262 (0.056)	6,215
GMB MICS 2018	-0.107 (0.036)	-0.070 (0.023)	0.046 (0.028)	0.047 (0.029)	-0.061 (0.053)	7,405
GMB DHS 2019	-0.139 (0.041)	-0.080 (0.022)	0.034 (0.031)	0.037 (0.034)	-0.105 (0.059)	6,549
GNB MICS 2014	-0.182 (0.032)	-0.116 (0.019)	0.021 (0.023)	0.024 (0.027)	-0.162 (0.041)	6,601
GNB MICS 2018	-0.159 (0.031)	-0.109 (0.020)	0.040 (0.025)	0.051 (0.032)	-0.119 (0.041)	7,378
GTM DHS 2014	-0.075 (0.013)	-0.063 (0.011)	-0.007 (0.009)	-0.012 (0.018)	-0.081 (0.016)	21,383
HND DHS 2011	-0.007 (0.013)	-0.006 (0.011)	0.013 (0.010)	0.023 (0.018)	0.006 (0.016)	21,361
HND MICS 2019	-0.035 (0.012)	-0.036 (0.012)	0.002 (0.009)	0.004 (0.017)	-0.033 (0.015)	20,668
HTI DHS 1994	-0.074 (0.032)	-0.063 (0.026)	0.010 (0.024)	0.017 (0.041)	-0.064 (0.039)	4,818
HTI DHS 2000	-0.082 (0.021)	-0.070 (0.018)	-0.004 (0.016)	-0.007 (0.028)	-0.086 (0.027)	9,588
HTI DHS 2005	-0.162 (0.020)	-0.136 (0.016)	-0.016 (0.014)	-0.029 (0.026)	-0.178 (0.024)	9,990
HTI DHS 2012	-0.117 (0.019)	-0.095 (0.014)	-0.010 (0.013)	-0.020 (0.024)	-0.127 (0.022)	13,176
IND DHS 2005	-0.090 (0.008)	-0.067 (0.006)	0.024 (0.006)	0.039 (0.010)	-0.066 (0.010)	109,032
IND DHS 2015	-0.117 (0.004)	-0.090 (0.003)	0.027 (0.003)	0.041 (0.005)	-0.091 (0.005)	601,507
IND DHS 2019	-0.047 (0.004)	-0.039 (0.003)	0.017 (0.003)	0.029 (0.005)	-0.030 (0.005)	636,696
KEN DHS 1993	-0.024 (0.015)	-0.034 (0.021)	0.027 (0.023)	0.031 (0.028)	0.002 (0.028)	7,948
KEN DHS 1998	-0.093 (0.020)	-0.092 (0.019)	0.007 (0.017)	0.012 (0.030)	-0.087 (0.026)	8,379
KEN DHS 2003	-0.032 (0.019)	-0.031 (0.019)	0.006 (0.016)	0.011 (0.031)	-0.026 (0.025)	8,559
KEN DHS 2008	-0.084 (0.018)	-0.088 (0.018)	0.022 (0.015)	0.046 (0.032)	-0.062 (0.023)	9,056
KEN DHS 2014	-0.091 (0.008)	-0.100 (0.009)	0.011 (0.008)	0.020 (0.015)	-0.079 (0.012)	36,418
KEN DHS 2022	-0.061 (0.008)	-0.068 (0.009)	0.012 (0.008)	0.020 (0.014)	-0.049 (0.011)	37,911
KGZ DHS 2012	-0.091 (0.017)	-0.092 (0.017)	0.068 (0.016)	0.109 (0.026)	-0.023 (0.021)	8,039
KHM DHS 2010	-0.085 (0.014)	-0.071 (0.012)	0.004 (0.013)	0.006 (0.018)	-0.081 (0.019)	15,667
KHM DHS 2014	-0.076 (0.014)	-0.072 (0.013)	0.010 (0.013)	0.015 (0.019)	-0.066 (0.018)	15,825
KIR MICS 2018	0.020 (0.038)	0.015 (0.028)	-0.009 (0.028)	-0.014 (0.040)	0.011 (0.048)	3,071

Table A4: Effect of man's questionnaire on number of men in the household

	Eligible men		Ineligible men		Total men	
Survey	Absolute	Relative	Absolute	Relative	Absolute	N
LAO MICS 2017	-0.052 (0.011)	-0.044 (0.009)	0.017 (0.009)	0.026 (0.015)	-0.036 (0.014)	22,287
LBR DHS 2013	-0.144 (0.019)	-0.135 (0.017)	-0.045 (0.018)	-0.062 (0.024)	-0.190 (0.026)	9,332
LBR DHS 2019	-0.060 (0.019)	-0.056 (0.018)	0.019 (0.017)	0.033 (0.030)	-0.041 (0.026)	9,068
LSO DHS 2004	-0.160 (0.019)	-0.168 (0.018)	0.111 (0.021)	0.135 (0.027)	-0.049 (0.028)	8,586
LSO DHS 2009	-0.169 (0.017)	-0.190 (0.017)	0.186 (0.021)	0.215 (0.027)	0.017 (0.026)	9,391
LSO DHS 2014	-0.180 (0.017)	-0.212 (0.017)	0.137 (0.019)	0.177 (0.027)	-0.042 (0.024)	9,402
LSO MICS 2018	-0.003 (0.020)	-0.003 (0.020)	0.005 (0.014)	0.011 (0.029)	0.002 (0.025)	8,847
MDA DHS 2005	-0.067 (0.015)	-0.078 (0.017)	0.030 (0.011)	0.086 (0.033)	-0.037 (0.017)	11,076
MDA MICS 2012	-0.036 (0.014)	-0.062 (0.023)	0.024 (0.011)	0.053 (0.024)	-0.012 (0.015)	11,353
MDG DHS 2003	-0.160 (0.020)	-0.142 (0.017)	0.081 (0.018)	0.174 (0.040)	-0.078 (0.026)	8,406
MDG DHS 2008	-0.095 (0.013)	-0.084 (0.011)	0.039 (0.011)	0.076 (0.023)	-0.056 (0.017)	17,847
MLI DHS 1995	-0.158 (0.021)	-0.139 (0.018)	0.042 (0.021)	0.063 (0.031)	-0.116 (0.030)	8,716
MLI DHS 2001	-0.065 (0.018)	-0.061 (0.017)	0.038 (0.016)	0.061 (0.026)	-0.026 (0.024)	12,320
MLI DHS 2006	-0.099 (0.018)	-0.084 (0.015)	0.076 (0.016)	0.126 (0.027)	-0.023 (0.024)	12,959
MLI DHS 2012	-0.198 (0.017)	-0.175 (0.014)	0.091 (0.018)	0.126 (0.027)	-0.108 (0.025)	10,105
MLI MICS 2015	-0.177 (0.025)	-0.109 (0.014)	0.091 (0.021)	0.081 (0.019)	-0.087 (0.036)	11,830
MLI DHS 2018	-0.170 (0.019)	-0.144 (0.015)	0.024 (0.018)	0.035 (0.027)	-0.146 (0.027)	9,510
MMR DHS 2015	-0.081 (0.015)	-0.089 (0.015)	0.001 (0.014)	0.001 (0.018)	-0.080 (0.019)	12,500
MNG MICS 2013	-0.061 (0.012)	-0.067 (0.012)	-0.012 (0.009)	-0.027 (0.022)	-0.072 (0.013)	14,805
MNG MICS 2018	-0.004 (0.012)	-0.004 (0.014)	-0.011 (0.010)	-0.023 (0.021)	-0.015 (0.014)	13,798
MOZ DHS 1997	-0.011 (0.019)	-0.011 (0.019)	0.027 (0.018)	0.045 (0.030)	0.017 (0.025)	9,279
MOZ DHS 2003	-0.035 (0.019)	-0.032 (0.017)	0.060 (0.016)	0.113 (0.030)	0.024 (0.024)	12,309
MRT MICS 2007	-0.045 (0.020)	-0.041 (0.018)	0.001 (0.017)	0.001 (0.023)	-0.045 (0.027)	10,359
MRT MICS 2015	-0.062 (0.018)	-0.061 (0.017)	0.029 (0.016)	0.041 (0.022)	-0.032 (0.025)	11,764
MWI DHS 1992	0.003 (0.019)	0.004 (0.025)	0.033 (0.031)	0.040 (0.037)	0.036 (0.035)	5,323
MWI DHS 2000	-0.038 (0.016)	-0.039 (0.016)	0.036 (0.014)	0.075 (0.030)	-0.001 (0.021)	14,210
MWI DHS 2004	-0.102 (0.014)	-0.109 (0.014)	0.018 (0.014)	0.035 (0.026)	-0.083 (0.019)	13,656
MWI MICS 2006	-0.038 (0.009)	-0.043 (0.010)	0.013 (0.009)	0.027 (0.018)	-0.025 (0.013)	30,542
MWI DHS 2010	-0.062 (0.011)	-0.064 (0.011)	0.043 (0.010)	0.077 (0.019)	-0.019 (0.015)	24,819
MWI MICS 2013	-0.066 (0.010)	-0.069 (0.010)	0.039 (0.010)	0.073 (0.019)	-0.026 (0.014)	26,713
MWI DHS 2015	-0.086 (0.011)	-0.087 (0.010)	-0.012 (0.010)	-0.022 (0.017)	-0.098 (0.014)	26,361
MWI MICS 2019	-0.047 (0.011)	-0.050 (0.011)	-0.000 (0.010)	-0.000 (0.018)	-0.048 (0.014)	25,419
NAM DHS 2000	0.012 (0.027)	0.011 (0.024)	0.035 (0.019)	0.067 (0.038)	0.047 (0.033)	6,380
NAM DHS 2006	-0.076 (0.022)	-0.073 (0.020)	-0.001 (0.017)	-0.002 (0.028)	-0.077 (0.028)	9,187
NAM DHS 2013	-0.040 (0.020)	-0.036 (0.018)	0.000 (0.013)	0.000 (0.034)	-0.040 (0.024)	9,842
NER DHS 1998	-0.102 (0.028)	-0.084 (0.022)	0.033 (0.025)	0.047 (0.037)	-0.069 (0.040)	5,927
NER DHS 2006	-0.158 (0.024)	-0.136 (0.019)	0.050 (0.021)	0.069 (0.030)	-0.108 (0.033)	7,654
NER DHS 2012	-0.173 (0.017)	-0.167 (0.015)	0.047 (0.018)	0.064 (0.025)	-0.126 (0.025)	10,747
NGA DHS 2003	-0.110 (0.025)	-0.091 (0.020)	0.033 (0.020)	0.063 (0.040)	-0.077 (0.033)	7,212
NGA DHS 2008	-0.085 (0.009)	-0.078 (0.008)	0.033 (0.008)	0.073 (0.018)	-0.053 (0.013)	34,023
NGA DHS 2013	-0.008 (0.008)	-0.009 (0.009)	0.004 (0.008)	0.006 (0.014)	-0.004 (0.012)	38,508
NGA DHS 2018	-0.064 (0.009)	-0.061 (0.008)	0.061 (0.008)	0.134 (0.018)	-0.003 (0.012)	40,427
NIC DHS 1998	-0.075 (0.021)	-0.057 (0.016)	0.022 (0.018)	0.034 (0.028)	-0.053 (0.028)	11,523
NPL DHS 2006	-0.045 (0.019)	-0.041 (0.017)	0.007 (0.016)	0.011 (0.027)	-0.038 (0.025)	8,707
NPL DHS 2011	-0.064 (0.017)	-0.061 (0.016)	-0.038 (0.014)	-0.064 (0.023)	-0.102 (0.021)	10,826
NPL DHS 2016	-0.079 (0.016)	-0.093 (0.018)	-0.036 (0.014)	-0.048 (0.018)	-0.115 (0.021)	11,040
NPL MICS 2019	0.004 (0.014)	0.004 (0.016)	0.019 (0.012)	0.030 (0.019)	0.023 (0.018)	12,653
PER DHS 1996	-0.044 (0.020)	-0.036 (0.016)	0.030 (0.016)	0.051 (0.028)	-0.014 (0.025)	28,119
PHL DHS 2003	-0.055 (0.018)	-0.044 (0.014)	0.013 (0.015)	0.021 (0.024)	-0.042 (0.022)	12,585
PNG DHS 2016	-0.080 (0.016)	-0.065 (0.013)	0.030 (0.013)	0.042 (0.019)	-0.049 (0.020)	16,001
RWA MICS 2000	-0.082 (0.018)	-0.082 (0.018)	0.010 (0.015)	0.021 (0.032)	-0.072 (0.024)	9,684
RWA DHS 2005	-0.071 (0.017)	-0.068 (0.016)	-0.015 (0.013)	-0.035 (0.029)	-0.086 (0.021)	10,270
RWA DHS 2010	0.016 (0.015)	0.016 (0.015)	-0.039 (0.011)	-0.092 (0.025)	-0.023 (0.019)	12,532
SEN DHS 2005	-0.128 (0.039)	-0.067 (0.020)	0.003 (0.029)	0.002 (0.028)	-0.125 (0.053)	7,411
SEN DHS 2010	-0.058 (0.038)	-0.030 (0.019)	0.083 (0.026)	0.082 (0.026)	0.025 (0.050)	7,902
SEN DHS 2014	-0.168 (0.050)	-0.087 (0.024)	0.026 (0.033)	0.026 (0.034)	-0.142 (0.066)	4,231
SEN DHS 2015	-0.060 (0.048)	-0.032 (0.025)	0.078 (0.031)	0.088 (0.037)	0.018 (0.063)	4,511
SEN DHS 2016	-0.153 (0.046)	-0.082 (0.023)	0.006 (0.031)	0.006 (0.033)	-0.147 (0.062)	4,437
SLE DHS 2008	-0.291 (0.022)	-0.230 (0.015)	0.162 (0.022)	0.230 (0.034)	-0.129 (0.030)	7,284
SLE DHS 2013	-0.129 (0.018)	-0.097 (0.013)	0.052 (0.014)	0.087 (0.025)	-0.077 (0.023)	12,620
SLE MICS 2017	0.004 (0.015)	0.004 (0.015)	0.004 (0.012)	0.008 (0.021)	0.008 (0.019)	15,308
SLE DHS 2019	-0.150 (0.017)	-0.119 (0.013)	0.058 (0.014)	0.099 (0.025)	-0.091 (0.023)	13,399
SUR MICS 2018	0.021 (0.021)	0.021 (0.021)	0.019 (0.015)	0.033 (0.027)	0.039 (0.025)	7,914
TCA MICS 2019	-0.003 (0.036)	-0.005 (0.062)	0.007 (0.031)	0.016 (0.073)	0.004 (0.041)	1,447
TCD DHS 1996	0.001 (0.028)	0.001 (0.026)	0.027 (0.023)	0.045 (0.040)	0.028 (0.039)	6,835
TCD DHS 2004	-0.017 (0.031)	-0.015 (0.027)	0.031 (0.025)	0.052 (0.042)	0.014 (0.042)	5,367
TCD MICS 2019	0.007 (0.014)	0.006 (0.014)	-0.025 (0.013)	-0.036 (0.019)	-0.018 (0.020)	18,967
TGO DHS 1998	-0.186 (0.024)	-0.145 (0.018)	0.020 (0.021)	0.028 (0.030)	-0.166 (0.034)	7,515
TGO MICS 2010	0.007 (0.024)	0.007 (0.024)	0.059 (0.023)	0.101 (0.040)	0.067 (0.034)	6,029
TGO DHS 2013	-0.098 (0.019)	-0.090 (0.016)	0.043 (0.016)	0.083 (0.032)	-0.055 (0.025)	9,548
THA MICS 2019	-0.009 (0.007)	-0.013 (0.011)	0.017 (0.007)	0.026 (0.011)	0.008 (0.009)	35,569
THA MICS 2022	-0.013 (0.008)	-0.019 (0.012)	0.006 (0.007)	0.009 (0.012)	-0.007 (0.010)	29,949
TLS DHS 2009	-0.074 (0.019)	-0.060 (0.015)	0.001 (0.017)	0.001 (0.020)	-0.073 (0.025)	11,462
TON MICS 2019	-0.007 (0.040)	-0.006 (0.035)	-0.012 (0.032)	-0.015 (0.040)	-0.018 (0.051)	2,498
TUN MICS 2018	0.004 (0.018)	0.004 (0.019)	0.006 (0.014)	0.009 (0.021)	0.010 (0.022)	11,224
TUV MICS 2019	-0.238 (0.096)	-0.156 (0.059)	0.005 (0.067)	0.006 (0.075)	-0.233 (0.122)	694
TZA DHS 1991	-0.026 (0.025)	-0.023 (0.022)	0.055 (0.022)	0.083 (0.034)	0.028 (0.034)	8,326
TZA DHS 1996	-0.074 (0.022)	-0.071 (0.020)	0.087 (0.019)	0.145 (0.034)	0.013 (0.029)	7,967
TZA DHS 2004	-0.010 (0.019)	-0.011 (0.020)	0.028 (0.019)	0.038 (0.026)	0.018 (0.027)	9,735
TZA DHS 2010	-0.065 (0.019)	-0.067 (0.019)	0.047 (0.019)	0.063 (0.027)	-0.019 (0.027)	9,623
TZA DHS 2015	0.002 (0.017)	0.002 (0.018)	0.032 (0.017)	0.045 (0.023)	0.034 (0.024)	12,563

Table A4: Effect of man's questionnaire on number of men in the household

Survey	Eligible men Absolute	Eligible men Relative	Ineligible men Absolute	Ineligible men Relative	Total men Absolute	N
TZA DHS 2022	-0.074 (0.014)	-0.084 (0.015)	0.022 (0.013)	0.032 (0.019)	-0.052 (0.019)	15,705
UGA DHS 1995	-0.069 (0.020)	-0.073 (0.020)	0.077 (0.020)	0.151 (0.041)	0.008 (0.028)	7,549
UGA DHS 2000	-0.041 (0.019)	-0.045 (0.020)	0.034 (0.020)	0.060 (0.036)	-0.006 (0.028)	7,876
UGA DHS 2006	-0.022 (0.018)	-0.023 (0.019)	-0.010 (0.018)	-0.016 (0.029)	-0.032 (0.025)	8,870
UGA DHS 2011	-0.112 (0.019)	-0.111 (0.018)	0.027 (0.018)	0.048 (0.032)	-0.085 (0.026)	9,033
UGA DHS 2016	-0.064 (0.012)	-0.068 (0.013)	-0.033 (0.011)	-0.060 (0.020)	-0.097 (0.017)	19,588
UKR DHS 2007	-0.061 (0.012)	-0.104 (0.019)	0.026 (0.010)	0.054 (0.021)	-0.035 (0.014)	13,368
UZB DHS 2002	-0.146 (0.036)	-0.095 (0.022)	0.057 (0.028)	0.083 (0.042)	-0.089 (0.043)	3,363
VNM MICS 2020	-0.028 (0.012)	-0.033 (0.014)	0.010 (0.010)	0.018 (0.019)	-0.018 (0.014)	13,359
WSM MICS 2019	-0.040 (0.046)	-0.027 (0.031)	0.040 (0.035)	0.043 (0.037)	0.001 (0.060)	3,196
XKX MICS 2013	-0.020 (0.032)	-0.014 (0.022)	0.035 (0.022)	0.042 (0.027)	0.014 (0.039)	4,127
XKX MICS 2019	-0.012 (0.026)	-0.009 (0.020)	0.011 (0.018)	0.014 (0.024)	-0.001 (0.031)	5,124
ZAF DHS 2016	-0.023 (0.017)	-0.024 (0.018)	-0.004 (0.011)	-0.012 (0.032)	-0.027 (0.020)	11,079
ZMB DHS 1996	-0.057 (0.026)	-0.049 (0.022)	0.082 (0.023)	0.130 (0.038)	0.025 (0.035)	7,286
ZMB DHS 2001	-0.156 (0.023)	-0.133 (0.018)	0.047 (0.021)	0.079 (0.036)	-0.109 (0.031)	7,123
ZWE DHS 1994	-0.052 (0.024)	-0.050 (0.023)	-0.012 (0.022)	-0.018 (0.033)	-0.064 (0.033)	5,983
ZWE DHS 1999	-0.063 (0.022)	-0.063 (0.021)	0.024 (0.020)	0.042 (0.035)	-0.039 (0.029)	6,369
ZWE MICS 2014	-0.057 (0.013)	-0.063 (0.014)	-0.002 (0.011)	-0.005 (0.023)	-0.060 (0.018)	15,686
ZWE MICS 2019	-0.030 (0.015)	-0.034 (0.017)	0.027 (0.013)	0.061 (0.029)	-0.003 (0.019)	11,091

Notes: Relative regression coefficients are computed as absolute regression coefficients over the control mean. Standard errors are displayed in parentheses.

Table A5: Effect of woman's questionnaire on number of women in the household

Survey	Absolute	Relative	N
GAB DHS 2019	-0.021 (0.008)	-0.091 (0.034)	11,781
GHA DHS 2008	-0.121 (0.016)	-0.121 (0.015)	11,778
NAM DHS 2013	-0.003 (0.008)	-0.015 (0.042)	9,849

Notes: Relative regression coefficients are computed as absolute regression coefficients over the control mean. Standard errors are displayed in parentheses.

Table A6: Bounds on missing women from survey-census comparison

Survey	Eligible women Absolute	Ineligible women Absolute	Lower bound Relative	Upper bound Relative	N
AGO DHS 2015	-0.044 (0.009)	0.093 (0.008)	-0.063 (0.006)	-0.126 (0.011)	539,065
BEN DHS 2001	-0.155 (0.014)	0.026 (0.012)	-0.069 (0.006)	-0.139 (0.012)	123,950
BEN DHS 2011	-0.279 (0.008)	0.074 (0.007)	-0.136 (0.004)	-0.273 (0.008)	194,670
BEN MICS 2014	-0.179 (0.010)	0.041 (0.009)	-0.083 (0.005)	-0.165 (0.009)	192,364
BFA DHS 1998	0.003 (0.017)	0.146 (0.016)	-0.050 (0.007)	-0.101 (0.015)	165,865
BFA MICS 2006	0.037 (0.020)	0.293 (0.020)	-0.091 (0.009)	-0.183 (0.017)	240,602
BOL DHS 1994	-0.030 (0.009)	0.065 (0.008)	-0.044 (0.006)	-0.088 (0.012)	150,516
BOL DHS 2003	-0.022 (0.008)	0.045 (0.006)	-0.032 (0.005)	-0.065 (0.010)	212,911
CIV DHS 1998	0.110 (0.036)	0.166 (0.027)	-0.021 (0.013)	-0.041 (0.027)	265,179
CMR DHS 2004	-0.114 (0.011)	0.069 (0.009)	-0.075 (0.005)	-0.149 (0.010)	345,535
CMR MICS 2006	-0.243 (0.012)	0.047 (0.010)	-0.115 (0.006)	-0.231 (0.012)	346,001
COL DHS 2005	0.079 (0.007)	0.061 (0.004)	0.008 (0.004)	0.016 (0.007)	1,071,921
CRI MICS 2011	0.021 (0.019)	0.055 (0.013)	-0.017 (0.013)	-0.034 (0.025)	126,620
CUB MICS 2010	0.015 (0.013)	0.071 (0.010)	-0.038 (0.012)	-0.076 (0.024)	376,454
CUB MICS 2014	-0.034 (0.013)	0.089 (0.010)	-0.082 (0.012)	-0.163 (0.024)	376,712
DOM MICS 2000	-0.049 (0.016)	0.022 (0.010)	-0.030 (0.008)	-0.060 (0.017)	204,663
DOM DHS 2002	-0.054 (0.007)	0.018 (0.006)	-0.035 (0.005)	-0.069 (0.009)	226,875
ETH DHS 2005	0.022 (0.008)	0.134 (0.008)	-0.053 (0.006)	-0.107 (0.012)	294,555
GHA DHS 1998	-0.391 (0.011)	-0.089 (0.010)	-0.123 (0.006)	-0.246 (0.011)	371,542
GHA DHS 2008	-0.275 (0.012)	-0.003 (0.010)	-0.119 (0.007)	-0.237 (0.013)	545,826
GIN DHS 2012	-0.441 (0.015)	0.150 (0.014)	-0.168 (0.005)	-0.336 (0.011)	153,397
GTM DHS 1995	-0.021 (0.011)	0.053 (0.009)	-0.030 (0.006)	-0.060 (0.012)	169,086
HTI DHS 2005	-0.146 (0.012)	0.070 (0.010)	-0.085 (0.006)	-0.169 (0.012)	185,841
IDN MICS 2000	0.003 (0.008)	0.062 (0.007)	-0.027 (0.005)	-0.053 (0.011)	5,062,004
IDN DHS 2012	-0.005 (0.005)	0.065 (0.004)	-0.033 (0.004)	-0.065 (0.007)	6,061,110
KEN DHS 1989	-0.126 (0.013)	0.193 (0.012)	-0.145 (0.008)	-0.291 (0.016)	222,621
KEN DHS 1998	-0.112 (0.010)	0.101 (0.009)	-0.097 (0.006)	-0.194 (0.012)	319,701
KEN DHS 2008	-0.063 (0.012)	0.067 (0.011)	-0.061 (0.007)	-0.123 (0.014)	892,539
KHM DHS 2000	-0.014 (0.008)	0.061 (0.008)	-0.029 (0.004)	-0.057 (0.009)	227,777
KHM DHS 2005	-0.056 (0.011)	0.021 (0.009)	-0.030 (0.005)	-0.060 (0.011)	34,945
KHM DHS 2010	-0.077 (0.008)	0.030 (0.007)	-0.042 (0.004)	-0.083 (0.009)	295,935
KHM DHS 2014	-0.112 (0.010)	0.041 (0.009)	-0.062 (0.006)	-0.124 (0.011)	44,172
KHM DHS 2021	-0.183 (0.007)	0.042 (0.006)	-0.097 (0.004)	-0.193 (0.009)	373,281
LAO MICS 2006	-0.127 (0.012)	0.049 (0.011)	-0.061 (0.006)	-0.121 (0.011)	100,760

Table A6: Bounds on missing women from survey-census comparison

Survey	Eligible women Absolute	Ineligible women Absolute	Lower bound Relative	Upper bound Relative	N
LAO MICS 2017	-0.278 (0.007)	0.014 (0.006)	-0.101 (0.003)	-0.202 (0.006)	140,210
LBR DHS 2007	-0.156 (0.015)	0.063 (0.012)	-0.085 (0.007)	-0.170 (0.014)	73,260
LBR DHS 2009	-0.125 (0.019)	0.102 (0.016)	-0.088 (0.009)	-0.176 (0.018)	70,625
LSO DHS 2004	-0.200 (0.011)	0.090 (0.010)	-0.135 (0.007)	-0.269 (0.014)	49,099
MAR DHS 1992	-0.006 (0.014)	0.080 (0.010)	-0.029 (0.005)	-0.057 (0.011)	226,740
MAR DHS 2003	0.062 (0.011)	0.039 (0.007)	0.008 (0.004)	0.015 (0.008)	292,297
MEX MICS 2015	-0.030 (0.016)	0.029 (0.012)	-0.028 (0.011)	-0.057 (0.022)	2,849,555
MMR DHS 2015	-0.133 (0.009)	0.054 (0.007)	-0.076 (0.005)	-0.151 (0.010)	1,092,036
MNG MICS 2010	-0.071 (0.009)	0.044 (0.007)	-0.054 (0.006)	-0.108 (0.012)	77,675
MOZ DHS 1997	-0.018 (0.019)	0.134 (0.014)	-0.072 (0.010)	-0.145 (0.021)	366,810
MOZ MICS 2008	-0.006 (0.008)	0.095 (0.008)	-0.049 (0.006)	-0.098 (0.011)	469,429
MOZ DHS 2009	-0.100 (0.010)	0.044 (0.010)	-0.061 (0.006)	-0.122 (0.012)	459,990
MWI DHS 1996	-0.110 (0.017)	0.092 (0.018)	-0.095 (0.012)	-0.190 (0.024)	227,107
MWI DHS 2000	-0.103 (0.007)	0.084 (0.007)	-0.089 (0.005)	-0.177 (0.010)	238,355
MWI MICS 2006	-0.160 (0.005)	0.067 (0.006)	-0.107 (0.004)	-0.215 (0.008)	311,089
MWI DHS 2010	-0.073 (0.006)	0.087 (0.006)	-0.076 (0.004)	-0.153 (0.008)	305,814
NAM DHS 2013	-0.095 (0.012)	0.002 (0.009)	-0.041 (0.006)	-0.082 (0.013)	102,477
NER DHS 2012	-0.375 (0.011)	0.000 (0.012)	-0.129 (0.005)	-0.258 (0.011)	34,672
NPL DHS 2011	-0.112 (0.010)	-0.013 (0.008)	-0.038 (0.005)	-0.076 (0.010)	676,334
PER DHS 1991	0.121 (0.009)	0.095 (0.007)	0.011 (0.005)	0.022 (0.010)	483,608
PER DHS 2007	-0.096 (0.006)	0.073 (0.004)	-0.077 (0.003)	-0.153 (0.007)	706,727
PER DHS 2009	-0.104 (0.007)	0.047 (0.005)	-0.069 (0.004)	-0.137 (0.008)	688,434
PHL DHS 2008	-0.000 (0.009)	0.058 (0.007)	-0.025 (0.005)	-0.049 (0.010)	2,058,873
PRY DHS 1990	0.015 (0.013)	0.057 (0.011)	-0.019 (0.007)	-0.038 (0.015)	90,914
RWA DHS 1992	0.050 (0.011)	0.060 (0.010)	-0.005 (0.007)	-0.010 (0.014)	154,753
RWA DHS 2000	-0.101 (0.009)	0.096 (0.009)	-0.084 (0.005)	-0.168 (0.011)	182,820
RWA MICS 2000	0.040 (0.014)	0.085 (0.013)	-0.019 (0.008)	-0.038 (0.016)	178,295
RWA DHS 2010	0.004 (0.007)	0.038 (0.006)	-0.015 (0.004)	-0.031 (0.009)	251,746
RWA DHS 2013	-0.021 (0.012)	0.034 (0.011)	-0.025 (0.007)	-0.051 (0.015)	244,050
RWA DHS 2014	-0.030 (0.007)	0.029 (0.007)	-0.027 (0.005)	-0.053 (0.009)	251,911
SEN DHS 2012	0.112 (0.037)	0.195 (0.024)	-0.021 (0.008)	-0.041 (0.017)	148,146
SEN DHS 2014	0.089 (0.039)	0.200 (0.024)	-0.028 (0.009)	-0.056 (0.018)	148,204
SEN DHS 2015	-0.024 (0.033)	0.189 (0.021)	-0.054 (0.008)	-0.107 (0.015)	148,480

Notes: Standard errors are displayed in parentheses.

Table A7: Bounds on missing men from survey-census comparison

Survey	Eligible men Absolute	Ineligible men Absolute	Lower bound Relative	Upper bound Relative	N
BEN DHS 2001	-0.223 (0.017)	0.068 (0.015)	-0.117 (0.009)	-0.234 (0.017)	121,150
BEN DHS 2011	-0.294 (0.012)	0.064 (0.010)	-0.142 (0.006)	-0.284 (0.012)	183,196
BEN MICS 2014	-0.212 (0.016)	0.033 (0.015)	-0.104 (0.009)	-0.207 (0.017)	183,064
BFA DHS 1998	-0.099 (0.020)	0.131 (0.020)	-0.090 (0.010)	-0.180 (0.021)	163,494
BOL DHS 2003	-0.059 (0.012)	0.040 (0.009)	-0.044 (0.007)	-0.088 (0.014)	200,818
CUB MICS 2014	-0.065 (0.017)	0.043 (0.014)	-0.072 (0.017)	-0.143 (0.033)	372,267
GHA DHS 1998	-0.482 (0.018)	-0.103 (0.014)	-0.147 (0.009)	-0.294 (0.017)	367,671
GHA DHS 2008	-0.356 (0.011)	-0.007 (0.009)	-0.150 (0.006)	-0.299 (0.011)	545,826
KEN DHS 1998	-0.162 (0.015)	0.066 (0.013)	-0.105 (0.009)	-0.210 (0.018)	315,578
KEN DHS 2008	-0.150 (0.017)	0.027 (0.014)	-0.083 (0.010)	-0.166 (0.019)	888,067
KHM DHS 2010	-0.125 (0.012)	-0.001 (0.009)	-0.051 (0.006)	-0.103 (0.012)	288,141
KHM DHS 2014	-0.170 (0.016)	0.012 (0.012)	-0.079 (0.009)	-0.158 (0.017)	33,982
LAO MICS 2017	-0.246 (0.010)	0.003 (0.008)	-0.090 (0.005)	-0.180 (0.009)	129,100
LSO DHS 2004	-0.242 (0.016)	0.073 (0.012)	-0.151 (0.009)	-0.303 (0.018)	44,809
MMR DHS 2015	-0.222 (0.012)	0.031 (0.010)	-0.117 (0.007)	-0.234 (0.015)	1,085,881
MOZ DHS 1997	-0.017 (0.027)	0.099 (0.023)	-0.059 (0.019)	-0.118 (0.038)	360,639
MWI DHS 2000	-0.105 (0.017)	0.052 (0.014)	-0.075 (0.011)	-0.150 (0.021)	227,898
MWI MICS 2006	-0.147 (0.010)	0.020 (0.010)	-0.083 (0.007)	-0.166 (0.014)	290,859
MWI DHS 2010	-0.093 (0.011)	0.087 (0.010)	-0.088 (0.007)	-0.175 (0.015)	289,819
NER DHS 2012	-0.610 (0.015)	0.069 (0.015)	-0.242 (0.007)	-0.483 (0.014)	29,101
RWA DHS 2000	-0.135 (0.015)	0.092 (0.013)	-0.112 (0.010)	-0.224 (0.020)	176,276
RWA DHS 2010	-0.007 (0.011)	0.022 (0.008)	-0.014 (0.006)	-0.029 (0.013)	242,213
SEN DHS 2014	-0.226 (0.055)	0.053 (0.026)	-0.070 (0.014)	-0.140 (0.028)	146,117
SEN DHS 2015	-0.227 (0.042)	0.083 (0.025)	-0.078 (0.011)	-0.156 (0.021)	146,223
SLE DHS 2013	-0.283 (0.016)	0.127 (0.012)	-0.139 (0.007)	-0.278 (0.013)	132,227
TGO MICS 2010	-0.032 (0.025)	0.059 (0.018)	-0.038 (0.012)	-0.077 (0.025)	121,355
TZA DHS 2004	-0.053 (0.019)	0.035 (0.017)	-0.044 (0.012)	-0.089 (0.024)	812,977
TZA DHS 2010	-0.103 (0.020)	0.058 (0.018)	-0.078 (0.012)	-0.157 (0.025)	942,302
UGA DHS 2000	-0.129 (0.017)	0.039 (0.017)	-0.083 (0.012)	-0.167 (0.023)	509,239
UGA DHS 2016	-0.090 (0.012)	0.009 (0.010)	-0.050 (0.008)	-0.101 (0.015)	717,523
VNM MICS 2020	-0.134 (0.010)	0.030 (0.009)	-0.089 (0.008)	-0.178 (0.016)	2,262,794

Notes: Standard errors are displayed in parentheses.

Table A8: Effect of man's questionnaire on the characteristics of eligible men

Survey	Age	Closely related to household head	Years of schooling	Ever married	Number of biological children in household	N
ALB DHS 2008	0.011 (0.008)	0.001 (0.003)	0.001 (0.007)	0.009 (0.022)	0.037 (0.037)	6,532
ALB DHS 2017	0.007 (0.005)	0.000 (0.002)	-0.001 (0.006)	0.008 (0.011)	-0.027 (0.032)	14,980
ARM DHS 2000	0.004 (0.007)	0.007 (0.003)	0.016 (0.006)		0.041 (0.040)	5,961
ARM DHS 2005	0.011 (0.009)	0.002 (0.003)	0.012 (0.010)		0.035 (0.046)	5,493
ARM DHS 2010	-0.011 (0.008)	-0.006 (0.004)	0.008 (0.006)	-0.021 (0.025)	0.025 (0.050)	5,224
ARM DHS 2015	0.004 (0.007)	-0.004 (0.003)	-0.010 (0.006)	0.011 (0.023)	0.016 (0.039)	5,786
AZE DHS 2006	0.001 (0.007)	0.004 (0.003)	0.008 (0.005)	0.018 (0.016)	0.104 (0.042)	8,641
BDI DHS 2010	0.014 (0.007)	-0.000 (0.008)	0.015 (0.018)	0.020 (0.018)	0.060 (0.028)	9,301
BDI DHS 2016	0.015 (0.005)	0.018 (0.006)	-0.004 (0.012)	0.042 (0.014)	0.055 (0.021)	16,360
BEN DHS 1996	0.002 (0.010)	0.019 (0.011)	0.021 (0.041)		0.051 (0.038)	4,339
BEN DHS 2001	0.013 (0.009)	0.013 (0.009)	-0.015 (0.026)		0.121 (0.039)	6,116
BEN DHS 2006	0.014 (0.005)	0.004 (0.005)	0.016 (0.015)	0.076 (0.013)	0.102 (0.021)	18,659
BEN DHS 2011	0.018 (0.006)	0.016 (0.005)	0.026 (0.015)	0.071 (0.012)	0.116 (0.022)	18,552
BEN MICS 2014	0.012 (0.006)	0.021 (0.006)	-0.010 (0.014)		0.070 (0.032)	14,559
BFA DHS 1998	0.019 (0.010)	0.008 (0.012)	0.001 (0.041)	0.047 (0.024)		6,110
BFA DHS 2003	0.021 (0.007)	0.013 (0.008)	0.085 (0.032)		0.089 (0.032)	12,275
BFA DHS 2010	0.028 (0.006)	0.014 (0.005)	0.014 (0.021)	0.077 (0.014)	0.120 (0.022)	16,286
BFA DHS 2021	0.030 (0.005)	0.021 (0.006)	-0.020 (0.015)	0.086 (0.014)	0.094 (0.022)	16,910
BGD DHS 2004	0.016 (0.005)	-0.002 (0.007)	-0.010 (0.015)	0.017 (0.013)		13,021
BOL DHS 1998	-0.001 (0.007)	0.000 (0.004)	0.003 (0.009)		-0.000 (0.028)	12,788
BOL DHS 2003	0.009 (0.005)	-0.004 (0.004)	-0.008 (0.006)		0.023 (0.023)	20,542
BOL DHS 2008	0.006 (0.005)	0.002 (0.003)	0.005 (0.006)		0.047 (0.024)	20,016
BRA DHS 1996	-0.009 (0.006)	0.005 (0.006)	0.027 (0.012)		-0.002 (0.032)	15,325
CAF DHS 1994	0.024 (0.010)	0.056 (0.015)	0.043 (0.026)		0.142 (0.051)	5,901
CAF MICS 2006	0.004 (0.006)	-0.000 (0.008)	0.025 (0.014)		0.080 (0.037)	11,028
CAF MICS 2010	0.001 (0.006)	0.003 (0.007)	0.007 (0.012)		0.115 (0.035)	11,175
CAF MICS 2018	-0.004 (0.007)	0.001 (0.010)	0.002 (0.012)		0.013 (0.036)	8,832
CIV DHS 1994	-0.021 (0.008)	0.031 (0.018)	0.029 (0.026)		0.013 (0.040)	8,700
CIV DHS 1998	0.014 (0.014)	0.010 (0.027)	0.077 (0.037)			3,120
CIV DHS 2011	0.013 (0.007)	0.012 (0.011)	0.021 (0.019)	0.056 (0.018)	0.157 (0.040)	11,852
CIV DHS 2021	-0.004 (0.005)	0.013 (0.007)	0.010 (0.015)	0.040 (0.014)	0.082 (0.030)	16,288
CMR DHS 1998	-0.009 (0.009)	0.022 (0.016)	0.020 (0.016)		-0.055 (0.044)	5,889
CMR MICS 2014	-0.000 (0.007)	0.009 (0.009)	0.026 (0.010)		-0.005 (0.040)	9,923
COD DHS 2007	0.018 (0.007)	0.014 (0.009)	0.015 (0.010)	0.068 (0.017)	0.088 (0.040)	10,575
COG DHS 2005	-0.003 (0.008)	-0.020 (0.012)	0.031 (0.012)		0.048 (0.042)	7,206
COG MICS 2014	-0.006 (0.006)	-0.012 (0.007)	0.012 (0.009)		-0.083 (0.034)	10,991
COM DHS 1996	-0.031 (0.015)	0.001 (0.025)	0.083 (0.049)		0.003 (0.065)	2,961
COM DHS 2012	-0.032 (0.010)	-0.027 (0.015)	0.001 (0.020)	-0.055 (0.025)	0.028 (0.050)	5,331
CUB MICS 2014	0.010 (0.006)	0.010 (0.006)	-0.010 (0.006)		0.112 (0.076)	7,190
CUB MICS 2019	0.000 (0.006)	0.004 (0.006)	0.011 (0.006)		0.058 (0.063)	7,757
ETH DHS 2000	0.020 (0.007)	-0.013 (0.009)	0.054 (0.023)		0.096 (0.040)	15,418
ETH DHS 2005	0.017 (0.006)	0.015 (0.006)	0.047 (0.015)		0.069 (0.029)	15,092
FJI MICS 2021	0.002 (0.008)	0.008 (0.010)	0.014 (0.008)		0.002 (0.043)	5,455
GAB DHS 2000	-0.013 (0.009)	0.014 (0.016)	0.016 (0.015)		0.049 (0.059)	7,303
GAB DHS 2012	0.002 (0.008)	0.006 (0.011)	0.002 (0.011)	0.064 (0.022)	0.139 (0.068)	9,210
GAB DHS 2019	-0.017 (0.007)	-0.025 (0.008)	0.023 (0.010)	-0.005 (0.018)	-0.023 (0.055)	11,442
GEO MICS 2018	0.007 (0.006)	0.011 (0.004)	0.009 (0.005)		0.115 (0.043)	8,877
GHA DHS 1998	0.005 (0.011)	-0.005 (0.008)	-0.032 (0.018)	0.055 (0.028)		4,867
GHA MICS 2006	0.007 (0.010)	-0.014 (0.009)	0.025 (0.016)		0.067 (0.054)	5,735
GHA DHS 2008	0.014 (0.007)	0.020 (0.006)	0.029 (0.011)	0.051 (0.018)		10,607
GHA MICS 2011	0.024 (0.008)	0.034 (0.006)	-0.003 (0.013)		-0.037 (0.044)	10,331
GHA DHS 2014	0.016 (0.007)	0.013 (0.006)	0.003 (0.011)	0.030 (0.019)		9,667
GHA MICS 2017	-0.002 (0.007)	-0.001 (0.006)	0.008 (0.009)		-0.030 (0.046)	11,096
GIN DHS 1999	0.015 (0.010)	-0.004 (0.013)	0.037 (0.038)		-0.004 (0.036)	7,038
GIN DHS 2005	0.021 (0.009)	0.015 (0.010)	0.018 (0.030)		0.106 (0.039)	7,031
GIN DHS 2018	0.041 (0.008)	-0.005 (0.009)	-0.002 (0.022)	0.087 (0.021)	0.116 (0.033)	9,213
GMB DHS 2013	-0.010 (0.008)	0.030 (0.016)	0.054 (0.021)	-0.012 (0.021)	-0.049 (0.042)	10,617
GMB MICS 2018	-0.014 (0.007)	0.003 (0.016)	0.023 (0.019)		-0.090 (0.042)	10,855
GMB DHS 2019	-0.004 (0.007)	0.013 (0.017)	0.061 (0.021)	-0.005 (0.020)	0.019 (0.041)	10,988
GNB MICS 2014	0.005 (0.007)	0.018 (0.014)	-0.015 (0.014)		0.034 (0.046)	9,784
GNB MICS 2018	0.007 (0.008)	0.002 (0.015)	0.015 (0.014)		0.021 (0.047)	10,415
GTM DHS 2014	0.003 (0.004)	-0.000 (0.003)	0.007 (0.008)	0.014 (0.010)	0.033 (0.020)	24,718
HND DHS 2011	0.007 (0.004)	-0.006 (0.004)	-0.005 (0.008)		0.014 (0.020)	25,326
HND MICS 2019	0.003 (0.005)	0.004 (0.004)	0.002 (0.007)		-0.004 (0.025)	19,674
HTI DHS 1994	-0.013 (0.010)	-0.020 (0.017)	-0.015 (0.026)		-0.027 (0.049)	5,568
HTI DHS 2000	-0.007 (0.008)	0.005 (0.011)	0.001 (0.017)		0.008 (0.057)	10,977
HTI DHS 2005	-0.004 (0.007)	0.013 (0.010)	-0.001 (0.015)	0.008 (0.018)	-0.012 (0.037)	11,093
HTI DHS 2012	-0.006 (0.006)	0.024 (0.009)	-0.001 (0.012)	0.013 (0.017)	0.060 (0.036)	15,135
IND DHS 2005	0.008 (0.002)	0.010 (0.003)	0.001 (0.004)	0.028 (0.005)	0.076 (0.013)	139,980
IND DHS 2015	0.008 (0.001)	0.005 (0.001)	-0.005 (0.002)	0.025 (0.003)	0.055 (0.007)	768,359
IND DHS 2019	0.002 (0.001)	0.002 (0.001)	0.004 (0.002)	0.008 (0.003)	0.019 (0.008)	766,282
KEN DHS 1993	0.020 (0.008)	0.015 (0.008)	-0.013 (0.013)		0.172 (0.043)	5,655
KEN DHS 1998	0.022 (0.008)	0.002 (0.008)	0.013 (0.010)		0.072 (0.041)	8,075
KEN DHS 2003	0.013 (0.007)	0.008 (0.008)	0.014 (0.011)		-0.022 (0.037)	8,600
KEN DHS 2008	0.006 (0.008)	0.009 (0.008)	-0.008 (0.009)	0.026 (0.021)		8,259
KEN DHS 2014	0.008 (0.004)	0.003 (0.004)	-0.002 (0.005)	0.025 (0.010)	0.013 (0.021)	31,482
KEN DHS 2022	-0.002 (0.004)	0.004 (0.004)	0.002 (0.004)	0.017 (0.011)	0.007 (0.022)	32,890
KGZ DHS 2012	0.010 (0.007)	-0.011 (0.005)	0.006 (0.005)	0.007 (0.017)	0.043 (0.034)	7,693
KHM DHS 2010	0.006 (0.005)	-0.009 (0.005)	0.000 (0.009)	0.030 (0.013)	0.015 (0.022)	18,018
KHM DHS 2014	0.017 (0.005)	0.003 (0.005)	0.013 (0.009)	0.040 (0.012)	0.058 (0.025)	16,461

Table A8: Effect of man's questionnaire on the characteristics of eligible men

Survey	Age	Closely related to household head	Years of schooling	Ever married	Number of biological children in household	N
KIR MICS 2018	-0.010 (0.009)	0.018 (0.017)	0.014 (0.010)		0.007 (0.043)	4,226
LAO MICS 2017	0.003 (0.003)	-0.001 (0.003)	-0.001 (0.006)		0.013 (0.016)	25,994
LBR DHS 2013	0.013 (0.007)	0.040 (0.012)	0.018 (0.014)	0.047 (0.018)	0.091 (0.045)	9,284
LBR DHS 2019	0.010 (0.008)	0.001 (0.010)	0.037 (0.015)	0.051 (0.018)	0.069 (0.046)	9,366
LSO DHS 2004	0.029 (0.010)	0.037 (0.010)	0.005 (0.017)		0.245 (0.060)	7,473
LSO DHS 2009	0.001 (0.009)	0.026 (0.010)	0.016 (0.015)	-0.012 (0.023)	0.040 (0.047)	7,502
LSO DHS 2014	0.003 (0.009)	0.003 (0.011)	-0.008 (0.014)	0.037 (0.025)	0.033 (0.053)	7,124
LSO MICS 2018	0.002 (0.007)	-0.008 (0.009)	0.013 (0.009)		-0.066 (0.047)	9,047
MDA DHS 2005	0.007 (0.007)	0.002 (0.003)	0.066 (0.054)		-0.039 (0.035)	9,252
MDA MICS 2012	0.003 (0.008)	0.009 (0.005)	0.008 (0.008)		-0.020 (0.044)	6,439
MDG DHS 2003	0.017 (0.008)	0.000 (0.007)	0.027 (0.014)		0.050 (0.044)	9,012
MDG DHS 2008	0.012 (0.005)	0.004 (0.005)	-0.019 (0.010)	0.030 (0.010)	0.035 (0.021)	19,338
MLI DHS 1995	0.022 (0.008)	0.004 (0.009)	-0.050 (0.037)		0.107 (0.032)	9,443
MLI DHS 2001	0.005 (0.007)	-0.022 (0.007)	-0.013 (0.031)		0.036 (0.031)	12,756
MLI DHS 2006	0.014 (0.006)	-0.004 (0.006)	0.044 (0.028)		0.031 (0.032)	14,743
MLI DHS 2012	0.040 (0.007)	0.018 (0.006)	-0.042 (0.024)	0.096 (0.016)	0.150 (0.028)	10,442
MLI MICS 2015	0.013 (0.005)	-0.005 (0.008)	0.014 (0.018)		0.030 (0.027)	18,184
MLI DHS 2018	0.029 (0.007)	0.034 (0.007)	0.003 (0.023)	0.096 (0.016)	0.146 (0.028)	10,431
MMR DHS 2015	0.005 (0.006)	0.013 (0.006)	-0.008 (0.010)	0.026 (0.015)	0.042 (0.029)	10,970
MNG MICS 2013	0.019 (0.005)	0.003 (0.003)	0.001 (0.007)		0.029 (0.022)	12,991
MNG MICS 2018	0.009 (0.005)	0.009 (0.003)	0.001 (0.008)		0.065 (0.031)	11,543
MOZ DHS 1997	0.014 (0.009)	-0.005 (0.009)	0.033 (0.017)		0.050 (0.072)	8,998
MOZ DHS 2003	0.007 (0.007)	-0.004 (0.008)	0.009 (0.016)		0.038 (0.031)	13,417
MRT MICS 2007	0.020 (0.007)	0.007 (0.009)	0.033 (0.021)		0.088 (0.039)	11,159
MRT MICS 2015	-0.018 (0.007)	0.000 (0.008)	0.002 (0.018)		-0.072 (0.038)	11,586
MWI DHS 1992	0.013 (0.010)	-0.001 (0.012)	0.008 (0.020)		0.059 (0.048)	4,003
MWI DHS 2000	0.008 (0.007)	0.009 (0.008)	-0.010 (0.011)		0.032 (0.031)	13,723
MWI DHS 2004	0.019 (0.007)	0.010 (0.007)	-0.033 (0.012)		0.062 (0.029)	12,234
MWI MICS 2006	0.011 (0.004)	0.005 (0.004)	-0.003 (0.007)		0.037 (0.022)	26,763
MWI DHS 2010	0.005 (0.005)	-0.001 (0.005)	0.001 (0.008)	0.009 (0.011)	0.011 (0.021)	23,558
MWI MICS 2013	0.007 (0.004)	0.008 (0.004)	0.003 (0.007)		-0.002 (0.023)	24,831
MWI DHS 2015	0.007 (0.005)	0.005 (0.005)	-0.002 (0.007)	0.054 (0.012)	0.031 (0.021)	25,285
MWI MICS 2019	-0.001 (0.005)	0.013 (0.004)	-0.003 (0.007)		-0.002 (0.024)	23,785
NAM DHS 2000	-0.010 (0.008)	-0.001 (0.015)	0.008 (0.014)		0.041 (0.066)	7,279
NAM DHS 2006	0.007 (0.007)	0.006 (0.013)	0.002 (0.011)	0.070 (0.034)	0.076 (0.060)	9,268
NAM DHS 2013	-0.005 (0.007)	0.006 (0.011)	-0.001 (0.010)	-0.001 (0.024)	-0.006 (0.048)	10,718
NER DHS 1998	0.028 (0.009)	0.015 (0.011)	0.035 (0.043)		0.122 (0.035)	6,849
NER DHS 2006	0.029 (0.008)	0.003 (0.009)	0.070 (0.036)		0.109 (0.034)	8,306
NER DHS 2012	0.038 (0.007)	0.002 (0.006)	0.039 (0.029)	0.089 (0.016)	0.143 (0.028)	10,242
NGA DHS 2003	0.002 (0.008)	-0.008 (0.009)	0.023 (0.015)		0.086 (0.047)	8,407
NGA DHS 2008	0.013 (0.004)	0.006 (0.003)	0.012 (0.006)	0.037 (0.010)	0.039 (0.016)	35,595
NGA DHS 2013	0.012 (0.004)	-0.009 (0.004)	0.016 (0.005)	0.009 (0.011)	0.052 (0.020)	35,801
NGA DHS 2018	0.019 (0.004)	0.004 (0.003)	0.019 (0.005)	0.116 (0.010)	0.219 (0.019)	41,909
NIC DHS 1998	-0.005 (0.007)	0.007 (0.006)	-0.001 (0.014)	0.056 (0.015)	0.057 (0.031)	14,975
NPL DHS 2006	-0.013 (0.007)	0.011 (0.007)	-0.010 (0.015)	-0.002 (0.012)	0.015 (0.028)	9,306
NPL DHS 2011	0.023 (0.007)	0.013 (0.007)	-0.007 (0.011)	0.020 (0.012)	0.001 (0.028)	11,022
NPL DHS 2016	0.005 (0.007)	0.014 (0.007)	0.006 (0.011)	0.029 (0.015)	0.062 (0.031)	8,902
NPL MICS 2019	-0.014 (0.005)	-0.004 (0.004)	0.024 (0.009)		-0.015 (0.026)	11,622
PER DHS 1996	0.000 (0.006)	-0.003 (0.007)	-0.009 (0.008)		0.108 (0.033)	34,583
PHL DHS 2003	0.003 (0.005)	0.009 (0.006)	-0.014 (0.007)			15,521
PNG DHS 2016	0.012 (0.005)	0.023 (0.007)	0.024 (0.009)	0.027 (0.014)	0.099 (0.033)	18,927
RWA MICS 2000	0.012 (0.008)	0.017 (0.009)	0.029 (0.017)		0.096 (0.035)	9,513
RWA DHS 2005	0.022 (0.007)	0.005 (0.007)	0.019 (0.016)	0.032 (0.019)	0.064 (0.029)	10,281
RWA DHS 2010	-0.000 (0.006)	0.004 (0.006)	-0.005 (0.012)	-0.007 (0.016)	-0.032 (0.023)	12,718
SEN DHS 2005	-0.005 (0.007)	0.002 (0.010)	0.070 (0.026)		0.030 (0.038)	13,845
SEN DHS 2010	-0.008 (0.006)	-0.011 (0.014)	0.020 (0.021)	-0.010 (0.017)	0.014 (0.036)	15,210
SEN DHS 2014	-0.001 (0.009)	0.006 (0.020)	0.021 (0.029)	-0.013 (0.024)	0.093 (0.057)	7,848
SEN DHS 2015	0.008 (0.009)	-0.023 (0.018)	-0.020 (0.027)	0.037 (0.025)	0.078 (0.048)	8,242
SEN DHS 2016	-0.002 (0.009)	-0.010 (0.017)	0.050 (0.027)	-0.028 (0.022)	0.007 (0.047)	7,995
SLE DHS 2008	0.022 (0.008)	0.052 (0.012)	0.020 (0.021)	0.092 (0.019)	0.181 (0.041)	8,137
SLE DHS 2013	0.003 (0.006)	-0.007 (0.009)	0.034 (0.015)	0.017 (0.013)	0.001 (0.024)	15,874
SLE MICS 2017	-0.012 (0.005)	-0.014 (0.008)	0.009 (0.014)		-0.029 (0.027)	15,041
SLE DHS 2019	0.017 (0.006)	0.018 (0.009)	0.025 (0.013)	0.053 (0.014)	0.113 (0.029)	15,832
SUR MICS 2018	0.004 (0.007)	-0.001 (0.007)	-0.009 (0.009)		-0.032 (0.047)	7,967
TCA MICS 2019	0.006 (0.020)	0.002 (0.023)	0.024 (0.017)		0.366 (0.284)	834
TCD DHS 1996	0.002 (0.009)	-0.044 (0.013)	0.038 (0.032)		0.074 (0.040)	7,398
TCD DHS 2004	0.013 (0.010)	-0.003 (0.013)	0.048 (0.033)		0.008 (0.044)	6,125
TCD MICS 2019	0.002 (0.005)	-0.007 (0.005)	0.020 (0.016)		0.019 (0.025)	19,619
TGO DHS 1998	0.026 (0.008)	0.021 (0.010)	0.012 (0.018)		0.147 (0.040)	8,899
TGO MICS 2010	0.008 (0.009)	-0.000 (0.010)	-0.016 (0.016)		-0.048 (0.040)	6,249
TGO DHS 2013	0.019 (0.007)	0.007 (0.008)	0.022 (0.013)	0.072 (0.019)	0.016 (0.032)	9,916
THA MICS 2019	-0.001 (0.004)	0.002 (0.003)	0.004 (0.005)		0.052 (0.040)	23,559
THA MICS 2022	0.000 (0.004)	-0.000 (0.003)	0.003 (0.005)		-0.038 (0.041)	19,874
TLS DHS 2009	0.008 (0.006)	0.003 (0.007)	0.011 (0.013)	0.029 (0.018)	0.025 (0.027)	13,804
TON MICS 2019	0.008 (0.013)	0.019 (0.015)	-0.001 (0.009)		-0.035 (0.069)	2,909
TUN MICS 2018	-0.008 (0.007)	0.005 (0.003)	-0.005 (0.009)		0.000 (0.037)	10,627
TUV MICS 2019	-0.026 (0.019)	0.041 (0.051)	0.002 (0.019)		-0.180 (0.100)	998
TZA DHS 1991	-0.003 (0.009)	-0.004 (0.010)	0.014 (0.014)		0.020 (0.043)	9,643
TZA DHS 1996	0.012 (0.009)	-0.020 (0.010)	0.047 (0.033)		0.038 (0.038)	8,088
TZA DHS 2004	0.000 (0.008)	-0.006 (0.009)	-0.021 (0.012)		0.029 (0.039)	9,065

Table A8: Effect of man's questionnaire on the characteristics of eligible men

Survey	Age	Closely related to household head	Years of schooling	Ever married	Number of biological children in household	N
TZA DHS 2010	0.002 (0.008)	-0.005 (0.009)	-0.006 (0.010)	0.005 (0.021)	-0.024 (0.037)	9,172
TZA DHS 2015	-0.005 (0.007)	-0.010 (0.008)	-0.008 (0.009)	-0.024 (0.018)	-0.016 (0.031)	11,995
TZA DHS 2022	0.005 (0.006)	0.013 (0.007)	-0.007 (0.009)	0.025 (0.017)	0.042 (0.031)	13,351
UGA DHS 1995	0.012 (0.009)	0.002 (0.010)	0.028 (0.015)		0.131 (0.043)	6,997
UGA DHS 2000	0.015 (0.009)	-0.005 (0.009)	0.012 (0.015)		0.094 (0.039)	7,074
UGA DHS 2006	0.015 (0.008)	-0.001 (0.009)	-0.009 (0.013)	0.053 (0.020)	0.039 (0.034)	8,257
UGA DHS 2011	0.013 (0.008)	0.002 (0.009)	0.001 (0.013)	0.062 (0.020)	0.070 (0.037)	8,742
UGA DHS 2016	0.008 (0.005)	0.009 (0.006)	0.027 (0.009)	0.027 (0.013)	0.035 (0.024)	17,929
UKR DHS 2007	0.003 (0.006)	0.003 (0.004)	-0.001 (0.005)		0.003 (0.040)	7,470
UZB DHS 2002	0.015 (0.008)	0.007 (0.005)	-0.024 (0.012)			4,981
VNM MICS 2020	-0.006 (0.005)	0.002 (0.003)	-0.002 (0.008)		0.024 (0.024)	11,009
WSM MICS 2019	-0.008 (0.010)	-0.003 (0.016)	0.004 (0.009)		0.023 (0.052)	4,637
XKX MICS 2013	-0.009 (0.008)	0.001 (0.007)	-0.006 (0.006)		-0.011 (0.037)	5,965
XKX MICS 2019	-0.001 (0.007)	0.001 (0.006)	0.002 (0.006)		0.046 (0.040)	6,452
ZAF DHS 2016	-0.012 (0.007)	-0.015 (0.008)	-0.010 (0.007)	-0.022 (0.025)	-0.060 (0.052)	10,142
ZMB DHS 1996	-0.022 (0.009)	0.015 (0.012)	0.004 (0.013)	-0.002 (0.021)	-0.016 (0.036)	8,401
ZMB DHS 2001	0.020 (0.008)	0.007 (0.011)	0.018 (0.011)		0.144 (0.037)	8,019
ZWE DHS 1994	0.003 (0.009)	-0.007 (0.012)	-0.019 (0.011)		0.033 (0.050)	5,993
ZWE DHS 1999	-0.012 (0.009)	-0.010 (0.011)	0.007 (0.010)		0.063 (0.051)	6,173
ZWE MICS 2014	0.004 (0.006)	0.015 (0.007)	-0.011 (0.005)		0.016 (0.028)	13,762
ZWE MICS 2019	0.001 (0.007)	0.016 (0.008)	0.003 (0.005)		-0.000 (0.032)	9,582

Notes: All regression coefficients are relative to the control mean. Standard errors are clustered at the household level and displayed in parentheses.

Table A9: Women's selection

Survey	Age	Closely related to household head	Years of schooling	Ever married	Children ever born	Survey	PHC
BEN DHS 2001	0.008 (0.004)	-0.115 (0.008)	0.040 (0.029)	0.010 (0.008)		6,448	154,594
BEN DHS 2011	0.034 (0.002)	-0.092 (0.005)	-0.093 (0.013)	0.055 (0.006)		17,329	229,892
BEN MICS 2014	0.043 (0.003)	-0.050 (0.007)	-0.056 (0.014)	0.054 (0.007)	0.182 (0.011)	16,348	237,416
BFA MICS 2006	0.031 (0.004)	0.015 (0.011)	0.177 (0.052)	0.006 (0.010)		8,159	329,415
BOL DHS 1994	0.004 (0.003)	-0.102 (0.007)	0.158 (0.011)	0.069 (0.009)		9,316	152,815
BOL DHS 2003	0.005 (0.003)	-0.159 (0.006)	0.024 (0.006)	0.116 (0.007)		18,487	200,216
CMR DHS 2004	-0.004 (0.003)	0.016 (0.007)	-0.067 (0.009)	0.215 (0.008)	0.119 (0.012)	11,304	412,147
CMR MICS 2006	0.014 (0.004)	-0.050 (0.008)	-0.021 (0.011)	0.110 (0.010)		9,408	422,494
CRI MICS 2011	0.002 (0.006)	0.018 (0.015)	0.000 (0.011)	0.154 (0.018)		5,740	121,704
CUB MICS 2010	-0.011 (0.005)	-0.025 (0.012)	0.072 (0.006)	0.174 (0.010)		9,440	276,307
CUB MICS 2014	0.011 (0.005)	-0.028 (0.011)	0.094 (0.006)	0.124 (0.011)		9,232	276,307
DOM MICS 2000	-0.015 (0.005)		0.015 (0.011)	-0.027 (0.011)		4,784	235,841
GHA DHS 1998	0.015 (0.004)	-0.245 (0.008)	0.145 (0.016)	0.057 (0.010)	0.043 (0.015)	4,970	449,300
GHA DHS 2008	0.012 (0.003)	-0.152 (0.006)	0.004 (0.008)	0.032 (0.008)	0.146 (0.018)	11,015	619,442
IDN MICS 2000	0.013 (0.003)	-0.046 (0.004)		-0.005 (0.006)		11,183	5,614,162
KEN DHS 1989	0.041 (0.005)	-0.083 (0.009)		0.104 (0.010)		7,424	236,014
KEN DHS 1998	0.025 (0.004)	-0.136 (0.008)	0.125 (0.009)	0.070 (0.009)	0.049 (0.012)	8,233	342,285
KEN DHS 2008	0.023 (0.005)	-0.005 (0.010)	0.083 (0.008)	0.043 (0.011)	0.049 (0.014)	8,767	934,904
KHM DHS 2000	0.020 (0.003)	-0.047 (0.006)	0.037 (0.011)	-0.006 (0.006)	0.034 (0.009)	15,557	281,213
KHM DHS 2010	0.024 (0.003)	0.003 (0.006)	0.080 (0.008)	0.052 (0.007)	0.092 (0.010)	19,237	358,486
KHM DHS 2014	0.017 (0.003)	0.032 (0.007)	0.009 (0.009)	0.125 (0.009)	0.163 (0.013)	18,012	34,975
KHM DHS 2021	0.027 (0.002)	-0.035 (0.006)	0.023 (0.007)	0.093 (0.006)	0.141 (0.009)	19,845	409,977
LAO MICS 2006	0.019 (0.004)	0.016 (0.006)	-0.095 (0.013)			7,703	137,057
LAO MICS 2017	0.034 (0.002)	-0.051 (0.003)	0.020 (0.007)	0.093 (0.005)	0.186 (0.007)	26,103	170,942
LBR DHS 2007	0.044 (0.005)	0.001 (0.011)	0.025 (0.021)	0.150 (0.011)	0.277 (0.017)	7,448	85,341
LBR DHS 2009	0.029 (0.006)	0.021 (0.014)			0.341 (0.021)	4,513	85,341
LSO DHS 2004	0.014 (0.004)	-0.020 (0.009)	-0.056 (0.006)	0.097 (0.011)	0.151 (0.016)	7,522	43,911
MEX MICS 2015	0.014 (0.005)	-0.063 (0.007)	0.007 (0.008)	0.089 (0.012)	0.108 (0.014)	12,937	2,989,055
MMR DHS 2015	0.027 (0.003)	-0.021 (0.007)	-0.023 (0.007)	0.025 (0.008)		13,454	1,341,553
MNG MICS 2010	0.032 (0.003)	-0.111 (0.007)		0.102 (0.009)		9,599	72,774
MOZ DHS 1997	0.017 (0.006)	0.025 (0.013)		0.059 (0.009)		9,590	377,199
MOZ MICS 2008	0.015 (0.003)	-0.012 (0.006)		0.062 (0.005)		15,060	472,585
MOZ DHS 2009	0.026 (0.006)	-0.040 (0.008)		0.074 (0.006)		6,749	534,121
MWI DHS 1996	0.034 (0.008)	-0.129 (0.011)	0.250 (0.025)	0.027 (0.011)	0.147 (0.024)	2,737	237,593
MWI DHS 2000	0.011 (0.003)	-0.094 (0.006)	0.084 (0.012)	0.026 (0.005)	0.024 (0.009)	13,538	237,593
MWI MICS 2006	-0.001 (0.002)	-0.081 (0.004)	-0.100 (0.007)	0.075 (0.004)	0.038 (0.007)	27,073	296,180
MWI DHS 2010	0.012 (0.002)	-0.044 (0.005)	0.049 (0.007)	0.001 (0.004)		23,748	295,369
NER DHS 2012	0.030 (0.004)	-0.217 (0.005)	-0.054 (0.029)	0.083 (0.005)		11,698	34,811
PER DHS 1991	-0.009 (0.002)	-0.013 (0.006)	0.052 (0.005)	-0.004 (0.007)	-0.031 (0.009)	17,351	570,535
PER DHS 2007	0.016 (0.002)	-0.073 (0.004)	0.054 (0.004)	0.005 (0.005)	0.041 (0.007)	42,636	730,539
PER DHS 2009	0.015 (0.002)	-0.091 (0.005)	-0.004 (0.004)	0.020 (0.007)	0.004 (0.008)	24,606	730,539
PRY DHS 1990	-0.004 (0.004)	-0.008 (0.010)	0.028 (0.009)	0.057 (0.011)	0.052 (0.015)	6,263	95,020
RWA DHS 1992	0.004 (0.004)	0.001 (0.008)		0.011 (0.009)		6,947	157,610
RWA DHS 2000	0.019 (0.003)	0.026 (0.008)	0.011 (0.011)	0.079 (0.008)		10,622	203,410
RWA MICS 2000	-0.010 (0.005)		-0.035 (0.016)	0.017 (0.013)		5,207	205,833
SEN DHS 2012	-0.025 (0.004)	0.056 (0.010)	-0.082 (0.025)	0.021 (0.011)		9,043	287,052
SEN DHS 2014	-0.004 (0.005)	0.067 (0.012)	0.033 (0.030)	0.030 (0.012)		8,831	287,052

Table A9: Women's selection

Survey	Age	Closely related to household head	Years of schooling	Ever married	Children ever born	Survey	PHC
SEN DHS 2015	-0.013 (0.004)	0.049 (0.010)	0.053 (0.027)	0.038 (0.011)		9,162	287,052
SLE DHS 2013	0.042 (0.003)	-0.003 (0.007)	-0.055 (0.015)	0.074 (0.007)	0.200 (0.011)	17,132	183,886
SLE DHS 2016	0.024 (0.004)	-0.021 (0.009)			0.309 (0.015)	8,526	183,886
TGO MICS 2010	0.019 (0.004)	-0.222 (0.008)	0.006 (0.017)	-0.004 (0.010)	0.048 (0.014)	7,016	143,932
TTO MICS 2011	0.012 (0.005)	-0.099 (0.009)	0.030 (0.006)	0.330 (0.022)		4,424	29,094
TZA DHS 2003	0.003 (0.004)	-0.098 (0.008)	0.069 (0.010)	0.044 (0.008)	-0.033 (0.011)	7,154	894,768
TZA DHS 2004	0.011 (0.003)	-0.115 (0.008)	0.026 (0.011)	0.066 (0.007)	-0.032 (0.010)	10,611	894,768
TZA DHS 2010	-0.004 (0.004)	-0.110 (0.008)	-0.043 (0.008)	0.110 (0.008)	-0.015 (0.011)	10,522	1,102,685
TZA DHS 2011	0.001 (0.004)	-0.089 (0.008)	0.015 (0.009)	0.093 (0.009)	-0.006 (0.010)	11,423	1,102,685
UGA DHS 2000	0.016 (0.004)	-0.066 (0.008)	-0.009 (0.012)	0.085 (0.008)	0.046 (0.011)	7,734	540,836
UGA DHS 2014	-0.003 (0.004)	-0.046 (0.010)			0.069 (0.014)	5,494	760,637
UGA DHS 2016	0.011 (0.002)	0.012 (0.007)	0.049 (0.006)	0.010 (0.005)	0.030 (0.008)	19,088	760,637
URY MICS 2012	0.023 (0.013)	-0.003 (0.046)		0.098 (0.043)		3,103	78,649
VEN MICS 2000	0.004 (0.004)	-0.015 (0.010)	0.055 (0.008)	-0.003 (0.012)		5,235	618,630
VNM MICS 2010	0.027 (0.003)	-0.079 (0.005)	-0.119 (0.005)	0.083 (0.006)	0.099 (0.009)	12,115	4,021,751
VNM MICS 2020	0.016 (0.003)	-0.122 (0.006)	0.014 (0.005)	0.096 (0.006)	0.178 (0.009)	11,294	2,077,336
ZAF DHS 2016	-0.007 (0.003)	0.007 (0.009)	0.017 (0.004)	-0.147 (0.014)		9,878	906,048
ZMB DHS 1992	-0.008 (0.004)	-0.054 (0.008)	0.104 (0.010)	0.100 (0.008)	0.161 (0.013)	7,250	177,735
ZMB DHS 2001	0.017 (0.003)	-0.038 (0.008)	0.093 (0.010)	0.045 (0.008)	0.108 (0.012)	7,944	217,666
ZWE DHS 2010	0.007 (0.003)	-0.026 (0.008)	-0.016 (0.004)	0.019 (0.007)	0.027 (0.010)	9,831	161,929

Notes: All regression coefficients are relative to the control mean. Standard errors are clustered at the household level and displayed in parentheses.

Table A10: Extreme rainfall events and question load across regression samples

	# eligible men					
	(1)	(2)	(3)	(4)	(5)	(6)
MQ	-0.091***	-0.096***	-0.095***	-0.084***	-0.088***	-0.088***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)
Drought	0.005	-0.018*	0.003			
	(0.011)	(0.010)	(0.011)			
Drought x MQ	0.030***	0.044***	0.033***			
	(0.008)	(0.009)	(0.008)			
Flood				0.005	0.011	0.004
				(0.010)	(0.012)	(0.010)
Flood x MQ				-0.018**	-0.016**	-0.015**
				(0.007)	(0.008)	(0.008)
Constant	1.034***	1.065***	1.090***	1.034***	1.061***	1.090***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Full sample	Yes	No	No	Yes	No	No
Married sample	No	Yes	No	No	Yes	No
Children sample	No	No	Yes	No	No	Yes
Number of surveys	73	47	63	73	47	63
Observations	865,214	648,684	786,581	865,214	648,684	786,581
R^2	0.109	0.063	0.065	0.109	0.063	0.065

All regressions include country-grid cell fixed effects and country-year fixed effects. MQ is an indicator variable that takes the value one if a household that is eligible for the man's questionnaire, and zero otherwise. Drought and flood events are defined as described in the text. Standard errors are clustered at the country-grid cell level.
* p<.10, ** p<.05, *** p<.01

Table A11: Man's questionnaire and missing men over time

	Length of man's questionnaire		Elasticity of sampled men		Share of missing men	
	(1)	(2)	(3)	(4)	(5)	(6)
2000s	63.9864***	57.9442***	-0.0022**	-0.0018	0.0196***	0.0181***
	(6.4742)	(6.9010)	(0.0009)	(0.0011)	(0.0050)	(0.0053)
2010s	75.7705***	69.0714***	-0.0014*	-0.0019**	0.0211***	0.0219***
	(5.6666)	(6.9697)	(0.0009)	(0.0010)	(0.0043)	(0.0053)
2020s	102.4518***	103.1996***	-0.0007	-0.0022	0.0219***	0.0278***
	(10.3637)	(8.9780)	(0.0014)	(0.0015)	(0.0067)	(0.0081)
Country FE	No	Yes	No	Yes	No	Yes
Mean 1990s	103.0357	103.0357	-0.0097	-0.0097	0.0612	0.0612
N	181	181	181	181	181	181
R^2	0.5152	0.7293	0.0265	0.4051	0.1558	0.5122

All specifications include survey program fixed effects. The omitted decade is the 1990s. The length of the man's questionnaire is measured by the number of questions listed in the questionnaire. See section 4.1.1 for details on the estimation of the share of missing men and section **??** for details on the estimation of the elasticity of sampled men. Robust standard errors in columns (1) and (2). Standard errors in columns (3)-(6) are bootstrapped using 100 repetitions.

Table A12: Elasticity of sample size and survey characteristics

	Dependent variable: Elasticity			
	(1)	(2)	(3)	(4)
Mandatory Re-interviewing	0.00437***			0.00588***
	(0.00149)			(0.00156)
Field Check Tables		0.00072		0.00022
		(0.00064)		(0.00059)
Tablet			-0.00019	-0.00222***
			(0.00080)	(0.00065)
Mean dep var	-0.01028	-0.01031	-0.01028	-0.01031
SD dep var	0.00952	0.00957	0.00952	0.00957
N	181	178	181	178
R^2	0.03766	0.02231	0.02096	0.04367

The dependent variable is the elasticity of sampled men with respect to question load. The independent variables *Mandatory re-interviewing*, *Field check tables* and *Tablet* are indicator variables that take the value one if a survey was implemented with the respective feature, and zero otherwise. Additional details are provided in appendix A.1.5. Standard errors are bootstrapped using 1000 repetitions.