Are We Done Yet?

Response Fatigue and Rural Livelihoods

Kate Ambler

Sylvan Herskowitz

Mywish Maredia

Markets, Trade, and Institutions Division

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI), a CGIAR Research Center established in 1975, provides research-based policy solutions to sustainably reduce poverty and end hunger and malnutrition. IFPRI's strategic research aims to foster a climate-resilient and sustainable food supply; promote healthy diets and nutrition for all; build inclusive and efficient markets, trade systems, and food industries; transform agricultural and rural economies; and strengthen institutions and governance. Gender is integrated in all the Institute's work. Partnerships, communications, capacity strengthening, and data and knowledge management are essential components to translate IFPRI's research from action to impact. The Institute's regional and country programs play a critical role in responding to demand for food policy research and in delivering holistic support for country-led development. IFPRI collaborates with partners around the world.

AUTHORS

Kate Ambler (k.ambler@cgiar.org) is a Research Fellow in the Markets, Trade, and Institutions Division of the International Food Policy Research Institute (IFPRI), Washington DC.

Sylvan Herskowitz (s.herskowitz@cgiar.org) is a Research Fellow in IFPRI's Markets, Trade, and Institutions Division, Washington, DC.

Mywish Maredia (maredia@msu.edu) is a Professor in the Department of Agricultural, Food, and Resource Economics at Michigan State University.

Are We Done Yet?
Response Fatigue and Rural Livelihoods

Kate Ambler†, Sylvan Herskowitz *†, and Mywish Maredia‡

†International Food Policy Research Institute
‡Michigan State University

December 8, 2020

Abstract

Accurate understanding of peoples' livelihoods activities is needed to inform effective policy. Existing evidence relies heavily on studies that use designated respondents to provide information about their household members, imposing significant costs on these respondents along with possible distortions in the data. In rural Ghana, we randomize the order that household members are asked about and estimate that response fatigue leads to undercounting of labor activities by 8% on average. Women are twice as impacted as men while youth are four times as impacted as older adults, distorting both within-household and population wide comparisons. These biases result from women and youth being listed systematically later in rosters and stronger effects of fatigue for them, conditional on roster position. The implications of our results extend to other topics of enquiry as well, wherever similar repetitive survey structures are deployed, such as birth records, plot-level inputs, and household consumption and expenditures.

Keywords: Labor, Response Fatigue, Gender, Youth, Survey Methodology
JEL Codes: O1, J2, C8, Q1

1

1 Introduction

Well-designed poverty alleviation programs and policies must be informed by an accurate understanding of peoples' livelihoods strategies and labor activities. In developing country settings, this understanding often relies on lengthy, multi-topic household surveys that document labor supply by collecting information on the productive activities of adult household members (Grosh et al., 2000). The recent policy focus on work opportunities for women and youth has driven a stream of studies characterizing the labor activities of these groups.[1] A related body of research has also explored the diversification of rural livelihood strategies, finding that off-farm activities are increasingly influential in the reduction of rural poverty.[2] The usefulness of this important work relies fundamentally on the quality of the available data.

In this paper, we examine an important and understudied factor affecting the quality of existing micro data on labor supply: response fatigue. Using random variation in the order of household members, we show that fewer unique labor activities are reported for household members listed later in the labor module, leading to underestimates of their labor contributions. We show that this underestimation differentially impacts women and youth, leading to economically meaningful and systematic biases in our understanding of peoples' livelihoods. We further show that non-randomized household rosters systematically list youth and women later, exacerbating losses in their reported labor activities.

This paper uses data from a survey in rural Ghana. Following a standard survey design, primary household respondents were asked to report details about their work activities as well as those of each household member age fourteen and above.[3] If the respondent indicated that a household member had worked over the past year, they were then asked a set of follow-up questions to identify the primary work activity and provide further details, including the type of work, amount of time spent on it, and earnings. They were then asked if that household member had a secondary activity and, if so, the

---

[1]For recent examples see Krumbiegel et al. (2020), Betcherman and Khan (2018), Van den Broeck and Kilic (2019), Bridges et al. (2011), and Klasen and Lamanna (2009).

[2]Recent examples include Dzanku (2020), Asfaw et al. (2019), Yeboah and Jayne (2018), Imai et al. (2015), Himanshu et al. (2013), Djurfeldt (2013), Haggblade et al. (2010) Ellis (1998), and Ellis and Freeman (2004).

[3]This is similar to the approach used in Living Standards Measurement Study surveys (Schaffner, 2000)

same set of follow-up questions describing that activity. It is almost immediately evident to respondents that by reporting fewer work activities they can reduce their response burden, shortening the number of questions they are asked and the time needed to complete the module. If respondents became "fatigued" from the repetitive questions, we would expect them to report fewer labor activities as they progressed through the list of eligible household members, thus underestimating the labor contributions of members listed later in the roster.

Instead of following the household order reported during the initial household listing, we randomize the order in which household members are asked about in the survey's labor module. This randomization enables us to measure the impact of response fatigue on reported labor supply activities resulting from being listed later in the sequence of household members. The median household in the sample has six eligible household members and we find a 2.2% reduction in the number of tasks reported for each spot a household member is listed later in the response order, controlling for household fixed effects and individual-level characteristics.

We also show that the losses from response fatigue differ for different groups. This bias comes from two distinct sources. First, the labor contributions of certain types of people may be systematically more vulnerable to fatigue-induced bias. This could be the case if their labor contributions are viewed as less important by the respondent and fatigue reduces their willingness to answer the follow-up questions required by each affirmative response. We test this directly in our data and find that the impacts of fatigue are stronger for both women and youth. A second possible source of bias is the endogenous ordering of household rosters. We show that women are positioned slightly later than men in the household roster, 0.2 positions on average, whereas youth age 14-24 are located 1.4 positions later on average than adults age 35-59. These systematic differences in roster position further contribute to the underestimation of labor activities for these groups.

We combine these two insights and use the reduced form models to generate estimates of the effects of response fatigue on reported labor activities for different sub-populations. We do this by generating "no fatigue" estimates of labor activities for everyone in the full sample. We then predict the number of work activities conducted for each person

after re-introducing losses from fatigue based on each household member's individual characteristics and the initial (non-random) household roster ordering. We do this first with a simple model that holds the impact of fatigue equal for all types of household members (but is still influenced by household roster position) and then for three versions of the model that allow for heterogeneity by age group, by gender, and in our preferred specification, by both. On average, we estimate that overall losses in terms of unreported labor activities are just under 8%. In our preferred model we find that these effects are nearly twice as large for women (10%) as for men (5%), and five times larger for youth age 14-24 (10-12%) than for adults age 35-59 (2.1%).

While our reported estimates are based on a rural sample in Ghana, our findings raise concerns that extend both outside this setting and beyond labor modules. Respondents, no matter their location, are just as likely to become tired or bored by similarly structured labor modules. This form of fatigue may also just as plausibly impact responses provided in survey modules on entirely different topics, such as household consumption, expenditures, agricultural production and input use, or birth histories, all of which typically follow iterative structures where affirmative responses for a given type of food, purchased item, farm input, or additional child trigger a set of follow-up questions. All of these modules are impacted by non-random ordering in the sequence of items or people and are therefore vulnerable to systematic biases. Certain types of modules may be even more vulnerable. For example, while the mean number of repetitions in this labor module was just below six, consumption modules frequently inquire about more than one hundred different food items, each with multiple follow-up questions, typically listing cereals first and sweets towards the end.[4]

Research on survey methodology has grown rapidly in recent years, propelled by the growth of household surveys in developing countries.[5] Topics addressed in this work have included the level of detail in questions, selection of the household's primary re-

---

[4]Beegle et al. (2012b) address the length of consumption modules by randomizing respondents into a long module, a short module with collapsed categories, or a short module focused on representative categories. They find that a short module with representative categories performs similarly to a long module, but they do not address the issue of data quality decay with the module.

[5]See for example the special issue on measurement in the Journal of Development Economics (McKenzie and Rosenzweig, 2012).

spondent, the use of proxy respondents, the use of screening questions, different recall windows, household definitions, and the use of high frequency checks.[6] Some papers have shown that design choices and features can induce biases in resulting data, consistent with the findings in this paper. For example, Galdo et al. (2020), Desiere and Costa (2019), and Comblon et al. (2015) show how question phrasing and recall windows can induce biases by age and gender in survey responses.

Despite this growing research on survey design, the issue of response fatigue has not received much attention in the labor and development literature. For example, in a review of lessons from 15 years of experience with the World Bank's Living Standards Measurement Study (LSMS) (Grosh et al., 2000), discussions on household roster construction (Glewwe, 2000) and labor modules (Schaffner, 2000) acknowledge concern that fatigue could lower data quality over the course of interviews, but do not study it directly nor acknowledge that it could be introducing systematic biases in the resulting data.

Response fatigue has, however, garnered some attention outside of labor and development economics. This includes research in contexts such as health and criminology that shows that data quality can deteriorate over the course of an interview (Bradley and Daly, 1994; Hess et al., 2012; Galesic and Bosnjak, 2009; Roberts et al., 2010; Egleston et al., 2011) and that attrition from future survey rounds is more likely when baseline surveys are longer (Rolstad et al., 2011; Hart et al., 2005). Additionally some articles have shown that the effects of response fatigue extend past adding noise to biasing the resulting data (Sharp and Frankel, 1983; Holbrook et al., 2007).

Our paper contributes to this literature by linking response fatigue to the understanding of rural livelihoods, showing statistically significant and economically meaningful losses in recorded labor activities. We also add to the literature by quantifying the losses for different sub-groups, and we demonstrate evidence of two distinct mechanisms lead-

---

[6]For examples of research on proxy respondents see Beegle et al. (2012b), Serneels et al. (2016), and Bardasi et al. (2011). Examples of question types and detail include Bardasi et al. (2011), Deininger et al. (2199), Langsten and Salen (2008), Comblon et al. (2015), and Benes and Walsh (2018). For examples of research on screening questions see Martin and Polivka (1995), Serneels et al. (2016), Dillon et al. (2012), and Fox and Pimhidzai (2013). For examples related to recall windows see Beegle et al. (2012a), Heath et al. (2020), Das et al. (2012), Deininger et al. (2199), Gaddis et al. (2020), and Arthi et al. (2016). See Beaman and Dillon (2012) for work on household definitions. For work on high frequency checks see Caeyers et al. (2012) and Fafchamps et al. (2012).

ing to these biases: the non-random ordering of household members in the roster and heterogeneity by age and gender in the estimated effect of response order. Our results speak directly, not just to the methodological literature on survey design, but also to research that bases its findings on characterizations of labor contributions estimated using conventional survey methodology.[7]

This paper proceeds by first providing background on the setting, data, and empirical strategy in Section 2. In Section 3 we provide our reduced form results. In Section 4 we model aggregate losses from fatigue. And in Section 5 we provide a brief discussion of results and recommendations, followed by concluding in Section 6.

2 Data and Empirical Strategy

We use data from a household survey conducted in Northern Ghana between April and June of 2019. The sample covered 12 districts in four regions. Respondents were members of farmer business organizations organized by the Ghana Agricultural Sector Investment Programme for the purpose of involvement in their agricultural programs. Sixty-six of these organizations, one per village, were included in this survey, serving as the baseline for a field experiment studying the adoption of conservation agriculture techniques. The survey included a detailed labor module in addition to others on agricultural activities, production practices, and knowledge.[8] Given the purpose of the original study, the survey targeted program participants as each household's primary respondent. In total, 1,106 households were interviewed as part of the study.

The survey's labor module was modeled after the Ghana LSMS survey Survey and the Uganda National Panel Survey. It followed a structure common to LSMS and other general surveys implemented by national statistics offices in developing countries. In the labor module, the respondent is asked about the labor activities of each household member age 14 and above. It begins by asking if the household member participated in any productive activities over the last 12 months. If so, they are asked a series of questions

---

[7]Notable examples include Dolislager et al. (2020), Yeboah and Jayne (2018), and Davis et al. (2010).
[8]See Ambler et al. (2020) for further details and results of the study.

to identify and describe their primary work activity over that period, how much time they devoted it, and their earnings. They are then asked if they had any secondary activities for that period and, if so, the follow-up questions are repeated. They are additionally asked to identify their primary productive activity over the last seven days, whether this was one of the activities already described, and if not, again asked to provide details about this third activity. Each respondent therefore has between 0 and 3 reported work activities.

This design is intended to capture details of the most important work activities for each individual over the past year, while also capturing primary activities over a seven day period to mitigate recall bias. However, respondents quickly learn from this structure that each answer they provide acknowledging a new work activity leads to a full set of follow-up questions. The median completion time for the labor module was 18.5 minutes, compared to a median completion time of 104 minutes for the full interview.[9] The median time spent per eligible household member was five minutes.[10]

Similar to many others, our survey uses proxy response for most observations so that the primary respondent reported about the activities of the members in their household. However, instructions in the survey allowed own response (or for the respondent to confer) if that member was readily available. In general, household surveys vary in the extent to which they allow for or prioritize proxy versus self reports, but our survey is common in employing an approach that allows for both own-reports and proxy reports. 17% of individuals in our analytical sample are recorded as having reported for themselves (or been conferred with), an incidence that is low, but within the range of other similar surveys.[11] Implications of this allowance for our results are discussed further in section 3.2.

We hypothesize that response fatigue increases as respondents progress through the survey and with each repetition of the labor module. As a result, respondents may report fewer distinct work activities for family members listed later in household rosters

---

[9]The labor module was the third section of the survey, and the median time elapsed prior to beginning the labor questions was 20.3 minutes.

[10]Conditional on the number of eligible household members, an additional member for which positive productive activities are reported increases total module length by approximately four minutes. The survey did not record the time spent on each individual, only the time spent on the entire module.

[11]The review by Desiere and Costa (2019) of the labor module data of LSMS surveys across six countries in Africa indicates that the rate of self-report responses ranged from 15% in Mali to 76% in Nigeria.

in order to avoid additional associated follow-up questions. Labor modules typically ask about household members in the order in which they were initially listed while creating the household roster. Similar to the guidance provided in most household surveys, the respondent in this survey was asked to list themselves first, followed by their spouse (if they had one), other adults in the house, and then children.[12] While these instructions are offered to ensure that household members are not unintentionally omitted, they also create systematic patterns across households in which roster order is mechanically related to gender and age. However, strict adherence to this ordering is not enforced (or observed in the data) and even with this guidance, respondents retain considerable discretion while filling out the roster. As such household roster orders may also capture biases or heuristics of the respondent such as listing higher earners or people with greater stature before others. Given the range of factors that likely influence the order of household listings in this surveys and many others, we cannot typically distinguish if correlations between work activities and listing order reflect real differences in household labor contributions or if they are driven or distorted by response fatigue.

In order to estimate the impact of fatigue and survey order on reported labor activities, we randomized the order of the household roster (excluding the respondent) in the labor module. Respondents reported information on their own labor activities first and then the module was repeated for each eligible member. Because the respondent's own position was not randomized, they are excluded from the analysis. Although they were included in the module, we additionally exclude those sixty years old and above (9% of the sample) from our analysis in order to focus on those who are below the retirement age and who are still capable of working.[13] Finally, our preferred specification uses household fixed effects and therefore requires at least two non-respondent household members. Households with fewer than two members aged 14 to 59 (excluding the respondent) are therefore

---

[12]Respondents were also told that a member of the household is someone who had slept in the respondent's house for 30 days consecutively or 60 days non-consecutively within the last 12 months, and shares food and other resources from a common source.

[13]Ghana's official retirement age is 60 years old and labor force participation drops sharply at this age. Additionally 44% of seniors are reported to be physically incapable of working whereas just 1.6% of household members below 60 are similarly incapacitated. We show robustness of our main results to including seniors in the analysis in Appendix Table A.2.

dropped from the analysis.[14]

2.1 Summary Statistics and "Balance"

Table 1 presents characteristics of the households and individuals included in the analysis sample. These households vary greatly in size but are large on average, with over eleven total members and six who were eligible for the labor module (14 or older). Approximately 30% were polygamous. 51% of respondents are female, 33% have ever attended school, and their average age is 42-years-old. The non-respondent household members in the analysis sample are 53% female, 26-years-old on average, and 56% are literate with 36% currently in school. The majority of household members are children (45%), spouses (29%), siblings (12%), or parents (4%) of the respondent.[15]

Household members are reported to have an average of 0.74 distinct work activities. 44% have no reported activities, 39% one activity, 18% two activities and just 0.3% had three distinct job activities listed. Among those who are working, 84% participate in household farm work as one of their activities listed. 15% participate in a household business and 25% engage in some form of hired, wage work. Excluding the primary respondent, the mean position in the labor module for household members is 3.8.

Column 1 of Table 2 shows that roster positions in household listings are strongly correlated with personal characteristics. The order variable is the household member's position among other members of the analysis sample, ranging from 1-14, where 1 represents the first member listed.[16] We present the partial correlations of each individual's gender, age group, student status, and relationship to the primary respondent with their order in the household roster, controlling for household fixed effects. Women are listed later by an average of 0.2 positions (p<.05). Younger individuals are also listed later in the roster. Individuals in the top age group, age 35-59, are listed an average of 1.4 positions

---

[14]Appendix Table A.1 compares samples resulting from these two selection criteria.

[15]Summary statistics of the full raw sample are shown in Appendix Table A.1. Characteristics are broadly similar but the analysis sample has larger households and is younger on average because the eligibility criteria drops people 60 and over and households with fewer than two people aged 14 to 59 other than the respondent.

[16]The top 1% of both the listed and randomized household orders are winsorized at position 14 to reduce the influence of a long right tail in the distribution of household sizes.

earlier than those aged 14-24 (p<.01), and those aged 25-34 are listed 0.4 positions earlier than the youngest group (p<.01). We also observe strong patterns with relationship to the respondent (spouse of the respondent is the omitted category). Finally, we note that those who report their own labor are listed 0.9 positions earlier than those who are not, and that student status is not significantly related to roster position.

In Column 2 of Table 2 we repeat the same exercise, but use the randomized order assignment as the dependent variable in place of the original listing order. Because of the randomization, we do not expect significant patterns and this analysis therefore serves as a balance check. The randomization appears to have been successful. All coefficients are small and only one (out of nine) is statistically significant at the 10% level.

2.2 Empirical Approach

We expect response fatigue to manifest in respondents reporting fewer work activities as the module is repeated for each household member. Our primary outcome of interest is "total jobs" defined as the number of distinct jobs listed during the module for that household member. Using the randomized ordering of household members in the labor module, we estimate the causal impact of response order on reported labor activities using the following regression specification:

$$jobs_{i,h} = \beta_0 + \beta_1 order_i + \beta_2 fem_i + \beta_3 inschool_i + \beta_4 self_i + \beta_5 Age14\_24_i + \beta_6 Age25\_34_i + \gamma_h + \psi_r + \epsilon_{i,h}$$

(1)

$jobs_{i,h}$ is the number of unique work activities listed for individual, i, from household, h, ranging from 0-3. $order_i$ is the individual's randomly assigned order number (1-14) for this individual.[17] fem_i indicates the gender of the individual, $inschool_i$ controls for whether the individual is currently in school. $self_i$ indicates if this household member was conferred with or reported for themselves during the labor module. $Age14\_24$ indicates that the individual was aged 14 to 24. $Age25\_34$ indicates the 25-34 age group, with age 35-59 as the omitted category. γ_h and ψ_r are a set of household and relation to respondent fixed effects. $\epsilon_{i,h}$ is the error term, clustered at the household level.

---

[17]The respondents themselves can be thought of as occupying position zero, before any effects of fatigue have accrued from repetition of the labor module.

3 Reduced Form Results

3.1 Main Effects and Heterogeneity

Table 3 presents estimates of the impact of response fatigue as induced by roster position on total jobs reported, building up to our preferred specification in equation (1). Column (1) shows regressions of the total number of jobs on response order, controlling only for fixed effects for the number of eligible household members in the labor module. Column (2) adds controls for gender, self-reporting, and student status. Column (3) adds controls for age cohort. Column (4) replaces the number of adults fixed effects with household fixed effects. Finally, Column (5) shows our preferred specification and includes a set of fixed effects for the household member's relationship to the respondent.

As expected given random assignment, the estimated impact of response order is nearly identical across columns. The preferred specification in column (5) indicates that individuals are reported to be involved in 0.016 (2.2%) fewer job activities for each position further back they are in the randomized ordering. This is an economically significant effect. An individual listed last in a household with five non-respondent adults in the labor module (the sample mean), would have their labor contributions under-reported by nearly 9% on average relative to the person listed immediately after the respondent. For an individual occupying the median listing position (3), labor activities are estimated to be under-reported by over 6.5%.

Table 4 presents a wide range of robustness checks, using the preferred specification in column (5) of Table 3. Column (1) reproduces the main result for reference. Columns (2) and (3) change the coding of the outcome to binary indicators for reporting at least one and at least two work activities, respectively. Point estimates of effects for these two outcomes are both highly significant (p<.01) and similar to one another at -0.007 and -0.008. However, this represents a 1.3% reduction in the likelihood of recording at least one activity whereas the likelihood of listing at least two declines by 4.6% per order position. This suggests that secondary work activities may be more vulnerable to fatigue than peoples' primary activities.

Next, we check robustness of our results to different transformations of the order vari-

11

able, to rule out that our results are driven only by the largest families.[18] Column (4) uses the logarithm of order position as the primary independent variable while column (5) uses the within household percentile location of each individual (scaled so that values range from zero to one). Both suggest consistently large, negative, and statistically significant impacts of later response position. Finally, we assess the sensitivity of the results to the choice of ordinary least squares for estimation using Tobit and Poisson models in columns (6) and (7) respectively. Results remain negative and highly significant. Additional robustness to the sample selection criteria is shown in Appendix Table A.2.

The results shown thus far assume that the relationship between order and reported outcomes is linear. This is an empirical question that we explore in Figure 1. In Figure 1 we show estimates of effects by response position, jointly estimated with the first listed member omitted as the reference group. Even for the second listed individual we see a negative effect relative to the first person listed of -0.05 ($p = 0.065$). The final coefficient, those who were tenth or later in their household listings, are reported to have 0.2 fewer distinct income earning activities.[19]

On average, the magnitudes of estimated negative effects increase with order. The p-value of the difference between the first and final positions is 0.06. While coefficients for individual order positions are noisily estimated and do not monotonically increase in magnitude, we consider the linear fit to be a reasonable first approximation, sufficient for the purposes of capturing the primary pattern of relevance: that response fatigue increases in order.

Next we explore whether the effects of response fatigue differ by the age and gender of household members in Table 5. Column (1) repeats the preferred result from column (5) of Table 3. Estimates in columns (2)-(4) show the effects separately by gender, age group, and gender-age group using fully saturated interactions on the order variable. The coefficients displayed therefore reflect estimation of the effect of fatigue for each subgroup and test whether this effect is significantly different from zero.

---

[18]The number of individuals in the labor module has a very long right tail. Our main results winsorize the top 1% of values, and these transformations employ two alternate approaches.

[19]We bundle those in positions ten or higher as the sample is small for these higher orderings. This final bin has approximately 5% of the sample.

Column (2) shows that the effect of response order on the reported labor activities of men is negative, but not significantly different from zero. By contrast, the effects of fatigue on reported labor activities of women are more than twice as large, over 3% of the mean (p<.01). A test for the difference between men and women (not shown) is statistically significant with p<.01. In column (3) we explore the effects of fatigue by different age groups. Effects for the oldest group are indistinguishable from zero, whereas they are highly significant for both of the younger age groups. While the point estimates are substantially larger for 25-34 year olds than for 14-24 year olds, given their different baseline levels of labor force participation, both represent losses of approximately 3% in reported labor activities per order position for the two groups. Finally, column (4) splits the sample by age and gender with the strongest impacts exhibited in the two younger groups of women. Young women appear to be doubly impacted by fatigue due to both their gender and age. The patterns shown are consistent with respondents valuing the labor contributions of youth and women less than those of older males, and indicate that the underestimation of labor activities caused by response fatigue is biased against these groups. In the next section we will build on the estimates in these four models to generate calculations of aggregate losses resulting from fatigue.

3.2   Further Analysis and Robustness

In this section we expand on several questions with potential to impact how we interpret and learn from our results. First, we examine the types of jobs that are being under-counted due to fatigue. Given that the results in Table 4 suggest that many jobs that are missed are secondary activities, it is reasonable to ask whether the jobs being missed are economically important. Appendix Table A.3 explores impacts of fatigue by type of work, reported pay, and the extensive and intensive margins of labor supply. The results suggest that wage work is disproportionately impacted relative to household farming and household-owned businesses. Though the coefficients on household farm work and wage work are similar in size, the lower overall levels of wage employment mean that the proportional effect is larger. The under-counting of these jobs could be especially detri-

mental to our ability to document off-farm work opportunities in rural areas that result from hired or seasonal work. We additionally note significant losses on the extensive margin of reporting receipt of any pay in the last week. Our results regarding the intensive margin of amount paid are noisily estimated and therefore less clear. Regarding hours worked, there is a strong negative impact on the extensive margin of working at all in the last week, but no statistically significant impact on number of hours worked. These results suggest that response fatigue may distort our understanding of the nature of rural employment, though we do not have sufficient power to make strong claims about the economic significance of the jobs that are impacted.

Second, average household size in our study sample is large, with an average of eleven household members and six members eligible for the labor module. As such, the external validity of our findings would be reduced if response fatigue was relevant only for large households. However, our analysis shows that losses from fatigue are still significant and of meaningful magnitude for people close to the beginning of the household ordering, as noted in the earlier discussion of Figure 1. Additionally, this does not account for any losses from fatigue that occur between the respondent answering about themselves and the first family member they are asked about.[20] Panel A of Appendix Figure A.1 repeats this exercise restricting the sample to only households with between five and seven members in the labor module. Again, significant negative effects are identified within a few roster positions of the respondent.

Appendix Table A.4 explores household size in a different way, splitting the sample into terciles by household size. Order effects are negative and highly significant even for households with just 2 to 4 members in the labor module with point estimates that are, in fact, bigger in magnitude than for larger households. This does not, however, mean that the aggregate impact of fatigue is just as big (or bigger) for small households. The point estimates reflect average marginal losses from being one position further back in a household roster. With more individuals in large households, and more repetitions of the labor module, aggregate losses from fatigue are still larger for these bigger households.

---

[20]Because respondents were always asked about first, we can not disentangle the effect of moving from the first to the second position in the order from the difference between self and other reporting.

The second three columns of Appendix Table A.4 show this by replacing the randomized order position variable with a variable that expresses that order as a percentile within the household. Reported coefficients therefore reflect the estimated difference between being randomly listed first versus last within households of different sizes. Column (4) shows that, though lacking in precision, the magnitude of estimated results for small households are economically meaningful, estimated to be 7% between the first and last position. However, these losses are much larger in large households with eight or more members in the labor module, showing losses of 17.9% between the first and last position (p<.01). Fatigue therefore results in larger losses in larger households, but the impacts of response order on our understanding of labor participation are important for households of all sizes.

Third, a potentially confounding factor is the use of self-reporting versus proxy response when completing the labor module. Other research has suggested that prioritizing self-reporting can reduce losses due to the difference between proxy versus self reporting (Benes and Walsh, 2018; Glewwe, 2000). It could be that this approach reduces losses from response fatigue as well. Our main specification includes an indicator for these instances of own-response as a control in our estimation.[21] Results are also robust to dropping these observations, shown in Appendix Table A.2. In Appendix Table A.6 we check for heterogeneity of our main effects by whether an individual responded for themselves. The estimated difference in response fatigue between proxy and self-reports is close to zero, though the standard error is large, limiting our ability to draw any strong conclusions. While more heavy reliance on self-reporting could, in theory, help to mitigate the effects of response fatigue, the process of locating all household members for response to the labor module is time consuming and costly and may not be feasible for many studies. Furthermore, encouragement or permission to seek self-reports has the potential of introducing different biases related to who is and who is not available for direct interrogation or conference.[22] Appendix Table A.6 looks at predictors of fatigue in our own sample as

---

[21]Own-response is not significantly less likely as the respondent progresses through the iterations of the labor module, shown in Table 2.

[22]The use of proxy respondents has been studied experimentally by (Bardasi et al., 2011) who find that proxy response does not affect the reporting of women's labor in Tanzania, but does lead to lower estimates of male employment.

well as those of three other household surveys and find that age and gender are strong predictors of whether an individual is available for self-reporting in their labor modules. Given that self-reporting is associated with higher reported numbers of activities, partial adherence to self-reporting could be unintentionally introducing greater biases by gender in age than the fatigue-induced biases it is intended to solve. The extent to which heavier reliance on self-reporting does or does not improve data quality remains a topic worthy of further research.

Finally, we consider the extent to which these results are attributable to respondent fatigue as opposed to fatigue experienced by the enumerator. Research has shown that enumerators can have substantial impacts on responses and data quality and therefore both channels are plausible (Di Maio and Fiala, 2199). Attribution of effects to respondent or enumerator could point to different solutions. Thus far, we have focused on the role of the respondent. This is because respondents control their answers to the enumerator's questions and how many activities they choose to report for each household member. By contrast, enumerators are trained to follow a strict script while conducting interviews. For enumerator fatigue to be driving our results, the enumerators would either need to be skipping questions and deviating from their scripts or, in some other way, be signalling that they want the respondent to under-report work activities. Additionally, to be consistent with the observed patterns of heterogeneity discussed in the analysis, enumerators would need to be differentially signalling their impatience across gender and age groups. Because the number of follow-up questions triggered by each additional work activity is the same regardless of an individual's characteristics, we would not expect enumerator fatigue to display the strong patterns of heterogeneity that were documented in Table 5.

We additionally test whether sensible patterns of enumerator fatigue can be detected in the data in Appendix Table A.7. First, we test if the estimated response fatigue varies by whether or not it was the enumerator's first survey of the day, hypothesizing that enumerators would be more patient at the beginning of the day.[23] We find that interviews conducted later in the day have, on average, slightly fewer reported work activities (not statistically significant) but do not find evidence of any differences in the relationship be-

---

[23]The modal number of surveys performed per day is two.

tween roster position and activities. We also categorize each survey by whether it was in the first or second half of the enumerator's total number of interviews. Surveys conducted in the second half have significantly fewer reported work activities, on average. However, again, we do not see any significant differential in the relationship between household roster position and reported activities by whether a survey was early or late in an enumerator's tenure. In results not shown we also rule out that enumerators who were better rated for performance react differently to response order. Ultimately, while we cannot rule out enumerator fatigue and recognize that it may still be affecting our estimates, our analysis suggests that this influence is likely to be secondary, and that our main results are driven predominantly by respondents.

4   Aggregate Losses from Fatigue

The estimates in the previous section show the impact of response fatigue on reported labor activities for marginal changes in roster position. However, fatigue-induced losses from one-position shifts do not capture aggregate losses and distortions that accrue through full administration of the labor module. In this section we use the reduced-form estimates from the previous section to estimate the aggregate effects of fatigue on recorded labor activities for the full population as well as to document biases induced by different impacts across sub-groups.

We begin by using the estimates from the previous section to predict what reported labor activities would be in the absence of response fatigue. After estimating a model, we set response order to zero for everyone in the sample, and use the resulting estimates to predict total jobs for each individual, preserving their baseline characteristics. This results in a set of predicted "no fatigue" labor activity levels for everyone in the sample. Setting response order to zero removes the effects of response fatigue from these predictions. We then generate a second set of predictions to model what labor activity levels would have been including the effects of response fatigue if the labor module had followed the non-randomized listing order generated during the creation of the household roster. To do this we replace the randomized listing order with the order from the initial household

listing and again predict individual activity levels. Comparing these two totals, we then calculate the predicted aggregate losses from fatigue for the full sample as well as for different subgroups.

Panel A of Table 6 presents results using the model based on equation (1) with a uniform fatigue parameter applied to all individuals as estimated in column (1) of Table 5. Panel B follows the same structure but shows results that allow for heterogeneity by age group and gender as estimated in Column (4) of Table 5. Appendix Table A.8 shows results for the intermediary models that allow separately for heterogeneity by age group and gender, as estimated in columns (2) and (3) in Table 5. In each panel we show the estimated means for the full sample, males, females, and people aged 14-24, 25-34, and 35-59. Column (1) of Table 6 shows the mean number of jobs for each sample from the raw data. Column (2) shows the predicted means without fatigue. Column (3) incorporates response fatigue with the endogenous initial household listing order. And Column (4) displays the percent losses for each group (between columns (2) and (3)).

In the first row of Panel A of Table 6 estimated losses are just under 8% for the full sample, similar to those estimated using the interacted model and shown in the first row of Panel B. The remaining rows present similar estimates and calculations for different sub-groups indicated in the first column of each row. The resulting percent losses by age group and gender for each of the four models are summarized in Figure 2.

We first examine the total losses to fatigue by age group in Panel A of Figure 2. Even though the simple model does not incorporate differential effects of fatigue by age, we still see sharp differences in the estimated losses by age group. This is driven by the strong ordering patterns shown in Table 2 whereby young people are systematically listed later in household rosters. Incorporating gender heterogeneity leads to similar results by age because gender is well balanced across age groups. The larger impacts of response order for individuals age 25-34, noted in Table 5, are reflected in results using the third and fourth models with age and age by gender heterogeneity, respectively. In the fully interacted model we see that both younger age brackets experience percent losses more than five times larger than those calculated for adults aged 35-59 with losses of 2%, 12%, and 10% for people aged 35-59, 25-34, and 14-24 respectively. This suggests that patterns

18

in the listing order that place youth systematically later in their household rosters as well as heterogeneity in the impact of response fatigue, conditional on their roster position, lead to younger individuals being disproportionately impacted and having their labor contributions under-counted.

Panel B of Figure 2 repeats the same sequence of models, but displays differences by gender instead of age group. Here, the simple model shows little differentiation between men and women resulting exclusively from listing order, consistent with women being listed just 0.2 positions later than men. However, allowing for heterogeneity in the effect of fatigue from response order by gender in the second model results in sharply different effects for men and women. In the fully interacted model, estimated losses in recorded job activities are 10% for women, nearly twice as large as those estimated for men, 5.3%.

5   Discussion

Our results show substantial losses from respondent fatigue in capturing labor force activities. These estimates are based solely on fatigue induced by random variation in response order within a survey module. If fatigue builds over the course of an entire survey, these results may present a lower bound on overall losses.

Our findings have implications for research that compares labor contributions across groups by either age or gender. While we cannot be sure that the same patterns of heterogeneity are present in other settings, roster listings from other commonly used household data sets follow similar demographic patterns. Table 7 shows that the 1993 Indonesian Family Life Survey, the 2012 Ghana GLS, the 2010 Nigerian LSMS, the 2012 Tanzanian LSMS, and the 2016 Malawi IHS all display strong patterns in the order in which household members are listed in the roster, also positioning women and youth systematically later in household rosters. In the presence of response fatigue in labor modules, these listing patterns alone will induce systematic biases, undercounting the contributions of women and youth in the labor market.

Employment opportunities for women and youth are a major current focus of researchers and policy makers. Providing productive employment opportunities for women

and young people is thought to be key for the continued development of rural areas. However, our results suggest that we may be systematically undercounting their contributions. Our analysis also indicates that secondary off-farm wage work activities are more vulnerable to undercounting due to response order; if true elsewhere, this could contribute to an understatement of the diversification of rural livelihoods. For example, Dolislager et al. (2020) provide a comprehensive accounting of youth employment across the world, and document strong sectoral and regional patterns. Given the focus on youth, the number and range of labor activities in that study may be underestimated and this underestimation may vary by sector. Where household size varies across regions, regional comparisons may also be biased by the role of response fatigue.

Having documented the potential harm of response fatigue, we now consider possibilities for ways to mitigate these distortions. First, researchers and data collection teams may want to consider avoiding the use of proxy and instead insisting on separate interviews and self-reporting. However, this may be a prohibitive approach for less well resourced studies, will not necessarily be effective (as discussed in section 3.2), and may unintentionally induce other dimensions of bias related to individuals' availability and willingness to participate.

A second possibility within the existing module structure and without cost implications is for researchers to consider their research objectives. If the primary interest is to make broad comparisons across groups or to better understand intra-household dynamics, randomization of the labor module ordering will mitigate the biases introduced by respondent fatigue (though not the average losses). If instead, capturing aggregate household labor activities and earnings is the priority, organizing the listing in order of greatest to least economic contribution would ensure that fatigue has the lowest impact on those whose contributions are most important to the household, minimizing losses.

A final possibility is to explore alternative survey designs. One approach could be to document the jobs that each individual does first, before iterating over each task. Initial activity listing could even be done for all household members first, before asking any follow-up questions. This could avoid some of the losses from fatigue, although it may come at the cost of a more disjointed interview experience and interfere with the ability to

encourage self-response from available household members. Ultimately, more methodological research is needed to understand the extent to which response fatigue impacts data quality across different contexts and topics. Fortunately, these experiments can be incorporated into existing studies at low cost.

6 Conclusion

Individual-level employment data and earnings statistics are key to the understanding of the structure of economic growth, the causes of poverty, and household welfare. In turn, this evidence base informs a wide range of policies and programs aiming to promote full and productive employment and decent work for all, a key concern of governments worldwide as evidenced by its inclusion in the Sustainable Development Goals. In this paper we explored response fatigue in the administration of household surveys as a source of measurement error and bias in estimates of reported labor activities.

The results suggest that response fatigue may be contributing to meaningful and widespread undercounting of rural labor activities. Average losses per individual are approximately 8% with especially large estimated losses for youth (10-12%) relative to older age groups (2%) and for women (10%) relative to men (5%). These differences are due both to heterogeneity in the effect of fatigue on respondents' reporting of activities for different groups of household members as well as the systematic manner in which women and young people are listed later in household rosters.

This type of response fatigue may extend past labor modules to other survey modules that follow repetitive structures. Topics include consumption and expenditure and agricultural production, suggesting that other important outcomes such as nutrition and household income could also be affected. Ultimately, improvement in survey design methods is key to our understanding of peoples' livelihoods strategies and well being. Without a reliable evidence base, policy and programs aiming to improve the lives of the world's poor will be unable to set reliable targets.

7 Main Tables and Figures

Figure 1: Effect of Randomized Order Position on Reported Total Job Activities
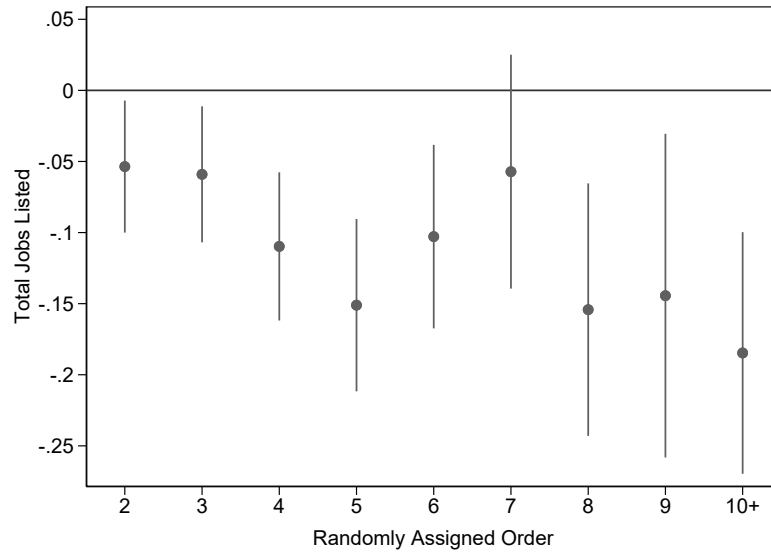
Notes: Figure shows the coefficients from jointly estimated random roster position assignments on total number of jobs listed (0-3). Estimation controls for household and relation to respondent fixed effects, gender, age group, schooling status and self-reporting.

Figure 2: Predicted Losses from Fatigue for Age and Gender Sub-Groups with Different Models

(a) Predicted Losses by Age Groups

(b) Predicted Losses by Gender
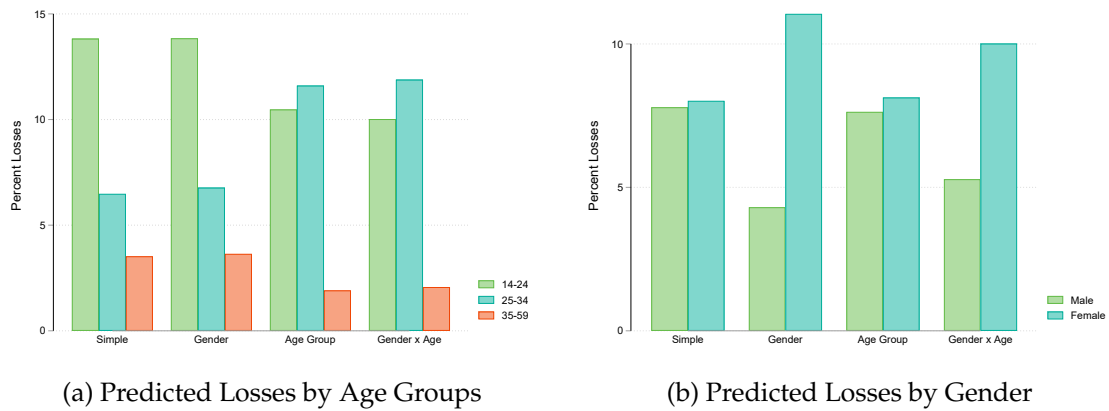
Notes: These figures show predicted aggregate losses from fatigue in reported number of labor activities for individuals by sub groups. Panel (a) shows differences by age group while Panel (b) shows differences by gender. The x-axes indicate the type of model used to predict fatigue. "Simple" uses a uniform fatigue factor regardless of the household member's characteristics. Gender indicates that estimates for fatigue are allowed to differ by gender. Age group, similarly, indicates different fatigue factors by age group. And the finale allows for differences in fatigue impacts by gender x age (six sub groups).

Table 1: Summary Statistics

Panel A: Household Level Characteristics	N=864	
	Mean	SD
Total Household Members	11.069	5.269
Members Included in Labor Module	5.918	2.904
Polygamous	.297	.457
Respondent: Female	.513	.5
Respondent: Age	42.294	11.715
Respondent: Ever School	.325	.469

Panel B: Household Member Characteristics	N=4252	
	Mean	SD
Female	.534	.499
Age	26.477	10.837
Literate	.564	.496
Currently in School	.356	.479
Self-Report on Labor	.171	.377
Relation to Respondent: Spouse	.189	.392
Relation to Respondent: Child	.447	.497
Relation to Respondent: Parent	.037	.189
Relation to Respondent: Sibling	.123	.328
Total Reported Job Activities	.744	.746
No Activities	.436	.496
One Activity	.386	.487
Two Activities	.175	.38
Three Activities	.003	.051
Household Farm Work among Working	.841	.366
Household Business among Working	.147	.354
Wage Work among Working	.253	.435
Labor Module Order Position	3.811	2.909

Notes: Panel A shows summary statistics at the household level including household and respondent characteristics. Panel B shows characteristics of the household members who participated in the labor module (above 14 and below 60 years old), excluding the respondent. Original household roster position is among those who are at least 14 years old and who were included in the labor module. Self-Report indicates those who reported their own labor activities or were otherwise conferred with during the survey.

Table 2: Order and Balance in Analysis Sample

	(1) Listed Order	(2) Randomized Order
Female	0.212**	-0.078
	(0.101)	(0.077)
Age 25-34	-0.413***	0.015
	(0.115)	(0.118)
Age 35-59	-1.359***	0.103
	(0.150)	(0.136)
Currently in School	0.121	-0.004
	(0.107)	(0.107)
Self-Report	-0.889***	-0.164
	(0.114)	(0.151)
Rel to Resp: Child	0.654***	-0.020
	(0.159)	(0.138)
Rel to Resp: Parent	0.534**	-0.106
	(0.261)	(0.275)
Rel to Resp: Sibling	0.913***	-0.312*
	(0.184)	(0.165)
Rel to Resp: Other	1.368***	0.153
	(0.184)	(0.149)
N	4252	4252
Mean Y	3.923	3.780
Resp Gender	All	All
Households	950	950
R2	0.576	0.436

Notes: Omitted age group is 14-24. Regressions include household fixed effects. Listed order recorded during initial household listing at beginning of survey. Randomized order implemented in sequence of household members in the labor module. Self-report indicates that household member reported for self or was conferred with during labor module.

Table 3: Fatigue and Randomized Order Position

	(1)	(2)	(3)	(4)	(5)
	Dependent Variable: Total Jobs Recorded				
Response Order (1-14)	-0.016***	-0.016***	-0.016***	-0.016***	-0.016***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Female		-0.066***	-0.057***	-0.067***	-0.077***
		(0.021)	(0.020)	(0.021)	(0.021)
Self-Report		0.095***	0.052*	0.095***	0.073**
		(0.029)	(0.027)	(0.031)	(0.031)
Student		-0.741***	-0.532***	-0.521***	-0.485***
		(0.023)	(0.028)	(0.028)	(0.028)
Age 25-34			0.258***	0.272***	0.224***
			(0.031)	(0.031)	(0.031)
Age 35-59			0.449***	0.454***	0.337***
			(0.032)	(0.034)	(0.041)
N	4252	4252	4252	4252	4252
Mean Y	0.744	0.744	0.744	0.744	0.744
Percent Effect	-2.085	-2.145	-2.202	-2.210	-2.183
Family Size FEs	Yes	Yes	Yes	No	No
Household FEs	No	No	No	Yes	Yes
Relation to Resp FEs	No	No	No	No	Yes
Households	950	950	950	950	950
R2	0.021	0.244	0.287	0.568	0.578

Notes: This table shows robustness of the estimated effects of randomized order position on reported number of jobs. Estimates in column (5) are the preferred specification from equation (1). Self-Report indicates that household member was conferred with during labor module.

Table 4: Robustness to Outcomes, Treatment Coding, and Estimation Method

	(1) Total Jobs	(2) One or More Jobs	(3) Two or More Jobs	(4) Total Jobs	(5) Total Jobs	(6) Total Jobs	(7) Total Jobs
Response Order (1-14)	-0.016***	-0.007***	-0.008***	-0.067***	-0.089***	-0.029***	-0.025***
	(0.004)	(0.003)	(0.003)	(0.015)	(0.024)	(0.007)	(0.006)
Female	-0.077***	-0.081***	0.001	-0.077***	-0.077***	-0.166***	-0.130***
	(0.021)	(0.014)	(0.012)	(0.020)	(0.021)	(0.036)	(0.033)
Self-Report	0.073**	0.042**	0.034*	0.074**	0.073**	0.116**	0.097**
	(0.031)	(0.019)	(0.019)	(0.031)	(0.031)	(0.052)	(0.049)
Student	-0.485***	-0.372***	-0.112***	-0.485***	-0.485***	-1.005***	-1.024***
	(0.028)	(0.021)	(0.015)	(0.028)	(0.028)	(0.056)	(0.069)
Age 25-34	0.224***	0.136***	0.087***	0.223***	0.222***	0.357***	0.303***
	(0.031)	(0.022)	(0.017)	(0.031)	(0.031)	(0.051)	(0.045)
Age 35-59	0.337***	0.211***	0.125***	0.337***	0.335***	0.495***	0.388***
	(0.041)	(0.026)	(0.024)	(0.041)	(0.041)	(0.064)	(0.055)
N	4252	4252	4252	4252	4252	4252	4252
Model	OLS	OLS	OLS	OLS	OLS	Tobit	Poisson
Order Variable	Level	Level	Level	Log	Percent	Level	Level
Mean Y	0.744	0.564	0.178	0.744	0.744	0.744	0.744
Percent Effect	-2.183	-1.293	-4.609	.	-11.927	-3.858	-3.305
Households	950	950	950	950	950	950	950
R2	0.578	0.584	0.434	0.578	0.577	0.353	0.216

Notes: "Level" indicates that the response order is a number from 1-14. "Log" takes the log of this variable. "Percent" is the within household position calculated so that the first person in the each household is set at zero and the final person per household set at 1. "Tot Jobs" is the total number of unique jobs listed for this individual. "One Plus" indicates a binary outcome for having at least one job listed in the labor module. "Two Plus" indicates at least two jobs listed. Regressions include household and relation to respondent fixed effects. Self-Report indicates that household member reported for self or was conferred with during labor module. The omitted age group is 14-24

	Dependent Variable: Total Jobs Recorded			
	(1)	(2)	(3)	(4)
Group 1: Order	-0.016***			
	(0.004)			
Group 1: Order x Male		-0.009		
		(0.005)		
Group 2: Order x Female		-0.023***		
		(0.005)		
Group 1: Order x Age 14-24			-0.012***	
			(0.004)	
Group 2: Order x Age 25-34			-0.031***	
			(0.008)	
Group 3: Order x Age 35-59			-0.009	
			(0.010)	
Group 1: Order x Male x Age 14-24				-0.007
				(0.005)
Group 2: Order x Male x Age 25-34				-0.014
				(0.009)
Group 3: Order x Male x Age 35-59				-0.009
				(0.012)
Group 4: Order x Female x Age 14-24				-0.017***
				(0.006)
Group 5: Order x Female x Age 25-34				-0.044***
				(0.009)
Group 6: Order x Female x Age 35-59				-0.009
				(0.011)
N	4252	4252	4252	4252
Group 1: Mean	0.744	0.738	0.450	0.482
Group 2: Mean		0.749	0.960	0.927
Group 3: Mean			1.204	1.192
Group 4: Mean				0.420
Group 5: Mean				0.984
Group 6: Mean				1.214
Group 1: Scaled Effect	-2.183	-1.173	-2.647	-1.408
Group 2: Scaled Effect		-3.084	-3.198	-1.531
Group 3: Scaled Effect			-0.715	-0.725
Group 4: Scaled Effect				-3.977
Group 5: Scaled Effect				-4.503
Group 6: Scaled Effect				-0.762
R2	0.578	0.578	0.578	0.580

Notes: The first column reproduces the main result from column (5) of Table 3. This table shows heterogeneity of the main effects of order position on total jobs listed. The estimates in each specification use a fully saturated treatment so that the main effect on each sub-group can be readily read from the table and tested against the null of no effect. In the statistics at the bottom of the table, group numbers indicate the relevant mean or scaled effect for the corresponding coefficients in the column above. For example, in column 3, "Group 1: Scaled Effect" suggests that youth age 14-24 have estimated losses per order position of 2.6%. "Group 5: Mean" in column 4 suggests that women between the age 25-34 have an average of 0.984 listed labor activities.

Table 6: Average Losses from Fatigue for Different Subgroups

Panel A: "Simple" Uniform Treatment Effect Model

	(1) Raw	(2) No Fatigue	(3) Fatigue	(4) % Diff
All	0.744	0.806	0.742	-7.912
	(0.746)	(0.565)	(0.574)	
Male	0.738	0.799	0.737	-7.791
	(0.726)	(0.561)	(0.570)	
Female	0.749	0.811	0.746	-8.015
	(0.763)	(0.568)	(0.577)	
Age 14-24	0.450	0.511	0.440	-13.824
	(0.649)	(0.474)	(0.477)	
Age 25-34	0.960	1.024	0.957	-6.482
	(0.737)	(0.472)	(0.479)	
Age 35-59	1.204	1.265	1.220	-3.524
	(0.652)	(0.423)	(0.430)	

Panel B: Interacted Heterogeneous Treatment Effect Model

	(1) Raw	(2) No Fatigue	(3) Fatigue	(4) % Diff
All	0.744	0.805	0.741	-7.896
	(0.746)	(0.576)	(0.574)	
Male	0.738	0.772	0.731	-5.283
	(0.726)	(0.534)	(0.554)	
Female	0.749	0.833	0.750	-10.01
	(0.763)	(0.609)	(0.591)	
Age 14-24	0.450	0.493	0.444	-10.02
	(0.649)	(0.477)	(0.478)	
Age 25-34	0.960	1.083	0.955	-11.883
	(0.737)	(0.480)	(0.490)	
Age 35-59	1.204	1.238	1.212	-2.067
	(0.652)	(0.432)	(0.431)	

Notes: Panel A uses a "simple" model to predict fatigue following equation (1): $totjobs_{i,h} = \beta_0 + \beta_1 order_i + \beta_2 female_i + \beta_3 inschool_i + \gamma_h + \psi_r + \delta_a + \epsilon_{i,h}$. Panel B uses a model allowing for heterogeneity of the impacts of fatigue by and age following the specification in Column (4) of Table 5.

The values in column (1) are means of the number of reported labor actvities in the raw data for individuals belonging to the group indicated at the start of each row. Column (2) removes predicted fatigue by setting $order_i = 0$ for all individuals then calculating the mean of these predicted "No Fatigue" values. Column (3) reimposes predicted fatigue if individuals had been asked about in the original, non-randomized, household listing order. Column (4) calculates the predicted proportion losses from fatigue.

Table 7: Predictors of Roster Order in Other Data Sources

	(1) Ghana GLS 6	(2) Nigeria LSMS 1	(3) Tanzania LSMS3	(4) Malawi IHS4	(5) Indonesia IFLS 1
Female	0.560***	0.813***	0.463***	0.388***	0.355***
	(0.017)	(0.033)	(0.027)	(0.016)	(0.022)
Age 25-34	-1.042***	-1.034***	-0.899***	-0.921***	-0.977***
	(0.029)	(0.054)	(0.042)	(0.032)	(0.036)
Age 35-59	-1.959***	-2.243***	-2.279***	-1.795***	-2.022***
	(0.023)	(0.049)	(0.033)	(0.024)	(0.027)
N	21354	11760	8757	11900	12178
Mean Y	2.707	3.107	2.863	2.498	2.718
R2	0.647	0.602	0.636	0.730	0.626

Notes: Dependent variable is household roster position among individuals age 14 and up. Omitted age group is youth age 14-24. Column headers indicate survey data sources. Regressions include household and relation to household head fixed effects.

References

Ambler, Kate, Alan de Brauw, and Mike Murphy (2020) "Increasing the adoption of conservation agriculture: A framed field experiment in Northern Ghana," URL: `https://ebrary.ifpri.org/digital/collection/p15738coll2/id/133739`.

Arthi, Vellore, Kathleen Beegle, Joachim De Weerdt, and Amparo Palacios-Lopez (2016) *Not your average job: Measuring farm labor in Tanzania*: The World Bank.

Asfaw, Solomon, Antonio Scognamillo, Gloria Di Caprera, Nicholas Sitko, and Adriana Ignaciuk (2019) "Heterogeneous impact of livelihood diversification on household welfare: Cross-country evidence from Sub-Saharan Africa," *World Development*, Vol. 117, pp. 278–295.

Bardasi, Elena, Kathleen Beegle, Andrew Dillon, and Pieter Serneels (2011) "Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania," *The World Bank Economic Review*, Vol. 25, pp. 418–447.

Beaman, Lori and Andrew Dillon (2012) "Do household definitions matter in survey design? Results from a randomized survey experiment in Mali," *Journal of Development Economics*, Vol. 98, pp. 124–135.

Beegle, Kathleen, Calogero Carletto, and Kristen Himelein (2012a) "Reliability of recall in agricultural data," *Journal of Development Economics*, Vol. 98, pp. 34–41.

Beegle, Kathleen, Joachim De Weerdt, Jed Friedman, and John Gibson (2012b) "Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania," *Journal of Development Economics*, Vol. 98, pp. 3–18.

Benes, Elisa and Kieran Walsh (2018) "Measuring employment in Labour Force Surveys: main findings from the ILO LFS pilot studies," *ILO*.

Betcherman, Gordon and Themrise Khan (2018) "Jobs for Africa's expanding youth cohort: a stocktaking of employment prospects and policy interventions," *IZA Journal of Development and Migration*, Vol. 8, p. 13.

Bradley, Mark and Andrew Daly (1994) "Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data," *Transportation*, Vol. 21, pp. 167–184.

Bridges, Sarah, David Lawson, and Sharifa Begum (2011) "Labour market outcomes in Bangladesh: The role of poverty and gender norms," *The European Journal of Development Research*, Vol. 23, pp. 459–487.

Van den Broeck, Goedele and Talip Kilic (2019) "Dynamics of off-farm employment in Sub-Saharan Africa: A gender perspective," *World Development*, Vol. 119, pp. 81–99.

Caeyers, Bet, Neil Chalmers, and Joachim De Weerdt (2012) "Improving consumption measurement and other survey data through CAPI: Evidence from a randomized experiment," *Journal of Development Economics*, Vol. 98, pp. 19–33.

Comblon, Virginie, Anne-Sophie Robilliard et al. (2015) "Are female employment statistics more sensitive than male ones to questionnaire design? Evidence from Cameroon, Mali and Senegal,"Technical report.

Das, Jishnu, Jeffrey Hammer, and Carolina Sánchez-Paramo (2012) "The impact of recall periods on reported morbidity and health seeking behavior," *Journal of Development Economics*, Vol. 98, pp. 76–88.

Davis, Benjamin, Paul Winters, Gero Carletto, Katia Covarrubias, Esteban J Quiñones, Alberto Zezza, Kostas Stamoulis, Carlo Azzarri, and Stefania DiGiuseppe (2010) "A cross-country comparison of rural income generating activities," *World development*, Vol. 38, pp. 48–63.

Deininger, Klaus, Calogero Carletto, Sara Savastano, and James Muwonge, "Can diaries help in improving agricultural production statistics? Evidence from Uganda," *Journal of Development Economics*, pp. 42–50.

Desiere, Sam and Valentina Costa (2019) *Employment Data in Household Surveys: Taking Stock, Looking Ahead*: The World Bank.

Di Maio, Michele and Nathan Fiala, "Be Wary of Those Who Ask: A Randomized Experiment on the Size and Determinants of the Enumerator Effect," *The World Bank Economic Review*, Vol. 34, pp. 654–669.

Dillon, Andrew, Elena Bardasi, Kathleen Beegle, and Pieter Serneels (2012) "Explaining Variation in Child Labor Statistics," *Journal of Development Economics*, Vol. 98, pp. 136–147.

Djurfeldt, Agnes Andersson (2013) "African re-agrarianization? Accumulation or pro-poor agricultural growth?" *World Development*, Vol. 41, pp. 217–231.

Dolislager, Michael, Thomas Reardon, Aslihan Arslan, Louise Fox, Saweda Liverpool-Tasie, Christine Sauer, and David L Tschirley (2020) "Youth and adult agrifood system employment in developing regions: Rural (peri-urban to hinterland) vs. urban," *The Journal of Development Studies*, pp. 1–23.

Dzanku, Fred Mawunyo (2020) "Poverty Reduction and Economic Livelihood Mobility in Rural Sub-Saharan Africa," *Journal of International Development*.

Egleston, Brian L, Suzanne M Miller, and Neal J Meropol (2011) "The impact of misclassification due to survey response fatigue on estimation and identifiability of treatment effects," *Statistics in medicine*, Vol. 30, pp. 3560–3572.

Ellis, Frank (1998) "Household strategies and rural livelihood diversification," *Journal of Development Studies*, Vol. 35, pp. 1–38.

Ellis, Frank and H Ade Freeman (2004) "Rural livelihoods and poverty reduction strategies in four African countries," *Journal of Development Studies*, Vol. 40, pp. 1–30.

Fafchamps, Marcel, David McKenzie, Simon Quinn, and Christopher Woodruff (2012) "Using PDA consistency checks to increase the precision of profits and sales measurement in panels," *Journal of Development Economics*, Vol. 98, pp. 51–57.

Fox, Louise and Obert Pimhidzai (2013) *Different Dreams, Same Bed: Collecting, Using, and Interpreting Employment Statistics in Sub-Saharan Africa-The Case of Uganda*: The World Bank.

Gaddis, Isis, Gbemisola Oseni, Amparo Palacios-Lopez, and Janneke Pieters (2020) "Measuring Farm Labor: Survey Experimental Evidence from Ghana," *The World Bank Economic Review*, DOI: `http://dx.doi.org/10.1093/wber/lhaa012`.

Galdo, Jose, Ana C Dammert, and Degnet Abebaw (2020) "Gender Bias in Agricultural Child Labor," *Working Paper*.

Galesic, Mirta and Michael Bosnjak (2009) "Effects of questionnaire length on participation and indicators of response quality in a web survey," *Public opinion quarterly*, Vol. 73, pp. 349–360.

Glewwe, Paul (2000) "Household Roster," in Margaret Grosh and Paul Glewwe eds. *Designing household survey questionnaires for developing countries*: The World Bank, Chap. 6, pp. 135–141.

Grosh, Margaret, Paul Glewwe et al. (2000) *Designing Household Survey Questionnaires for Developing Countries*: The World Bank.

Haggblade, Steven, Peter Hazell, and Thomas Reardon (2010) "The rural non-farm economy: Prospects for growth and poverty reduction," *World Development*, Vol. 38, pp. 1429–1441.

Hart, Timothy C, Callie Marie Rennison, and Chris Gibson (2005) "Revisiting respondent "fatigue bias" in the National Crime Victimization Survey," *Journal of Quantitative Criminology*, Vol. 21, pp. 345–363.

Heath, Rachel, Ghazala Mansuri, Bob Rijkers, William Hutchins Seitz, and Dhiraj Sharma (2020) "Measuring Employment: Experimental Evidence from Urban Ghana," *World Bank Policy Research Working Paper*.

Hess, Stephane, David A Hensher, and Andrew Daly (2012) "Not bored yet–revisiting respondent fatigue in stated choice experiments," *Transportation research part A: policy and practice*, Vol. 46, pp. 626–644.

Himanshu, Peter Lanjouw, Rinku Murgai, and Nicholas Stern (2013) *Non-farm diversification, poverty, economic mobility and income inequality: A case study in village India*: The World Bank.

Holbrook, Allyson L., Jon A. Krosnick, David Moore, and Roger Tourangeau (2007) "Response Order Effects in Dichotomous Categorical Questions Presented Orally: The Impact of Question and Respondent Attributes," *Public Opinion Quarterly*, Vol. 71, pp. 325–348.

Imai, Katsushi S, Raghav Gaiha, and Ganesh Thapa (2015) "Does non-farm sector employment reduce rural poverty and vulnerability? Evidence from Vietnam and India," *Journal of Asian Economics*, Vol. 36, pp. 47–61.

Klasen, Stephan and Francesca Lamanna (2009) "The impact of gender inequality in education and employment on economic growth: new evidence for a panel of countries," *Feminist Economics*, Vol. 15, pp. 91–132.

Krumbiegel, Katharina, Miet Maertens, and Meike Wollni (2020) "Can employment empower women? Female workers in the pineapple sector in Ghana," *Journal of Rural Studies*.

Langsten, Ray and Rania Salen (2008) "Two approaches to measuring women's work in developing countries: A comparison of survey data from Egypt," *Population and Development Review*, Vol. 34, pp. 283–305.

Martin, Elizabeth and Anne E Polivka (1995) "Diagnostics for redesigning survey questionnaires: Measuring work in the Current Population Survey," *Public Opinion Quarterly*, Vol. 59, pp. 547–567.

McKenzie, David and Mark Rosenzweig (2012) "Preface for symposium on measurement and survey design," *Journal of Development Economics*, Vol. 98, pp. 1–2.

Roberts, Caroline, Gillian Eva, Nick Allum, and Peter Lynn (2010) "Data quality in telephone surveys and the effect of questionnaire length: A cross-national experiment,"Technical report, ISER Working Paper Series.

Rolstad, Sindre, John Adler, and Anna Rydén (2011) "Response burden and questionnaire length: is shorter better? A review and meta-analysis," *Value in Health*, Vol. 14, pp. 1101–1108.

Schaffner, Julie Anderson (2000) "Employment," in Margaret Grosh and Paul Glewwe eds. *Designing household survey questionnaires for developing countries*: The World Bank, Chap. 9, pp. 135–141.

Serneels, Pieter, Kathleen Beegle, and Andrew Dillon (2016) *Do Returns to Education Depend on How and Who You Ask?*, Policy Research Working Papers: The World Bank, DOI: `http://dx.doi.org/10.1596/1813-9450-7747`.

Sharp, Laure M and Joanne Frankel (1983) "Respondent burden: A test of some common assumptions," *Public Opinion Quarterly*, Vol. 47, pp. 36–53.

Yeboah, Felix Kwame and Thomas S Jayne (2018) "Africa's evolving employment trends," *The Journal of Development Studies*, Vol. 54, pp. 803–832.
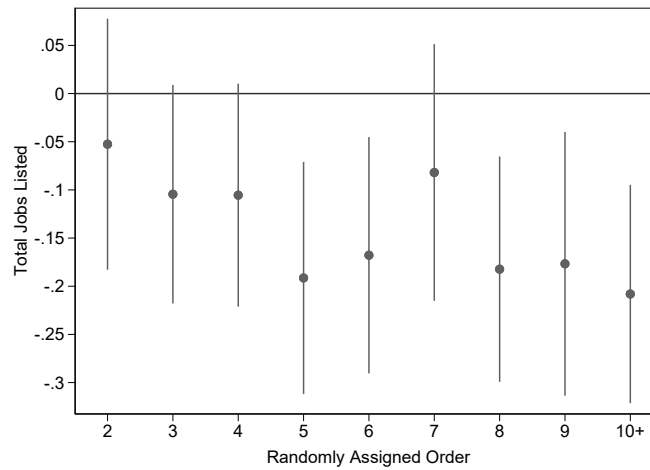
8 Appendix

Figure A.1: Ordering Effects on Subsets of Household Size

(a) Households with 5-7 People in Labor Module
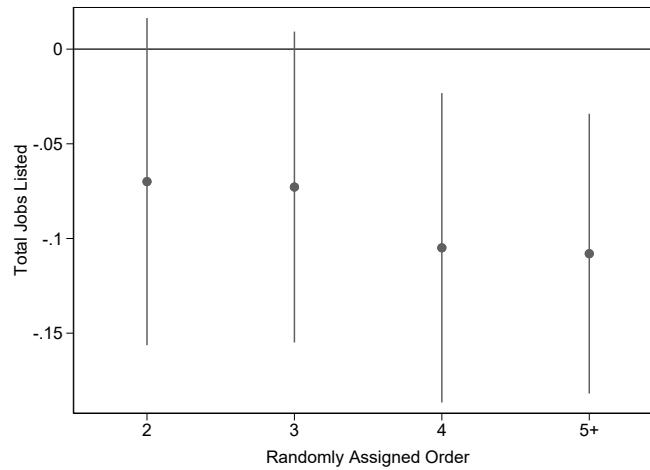
(b) Households with 8+ People in Labor Module

Notes: Individual effects by labor module order position for household size subgroups.

Table A.1: Summary Statistics for Raw and Analysis Samples

Sample:	Full/Raw Sample		3+ Eligible		3+ Eligible No Seniors	
Panel A: Household Level Characteristics	N=1106		N=988		N=864	
	Mean	SD	Mean	SD	Mean	SD
Total Household Members	10.476	5.234	11.012	5.261	11.069	5.269
Members Included in Labor Module	5.462	2.987	5.876	2.895	5.918	2.904
Polygamous	.276	.447	.301	.459	.297	.457
Respondent: Female	.517	.5	.515	.5	.513	.5
Respondent: Age	41.188	11.701	41.893	11.74	42.294	11.715
Respondent: Ever School	.349	.477	.332	.471	.325	.469

Sample:	Full/Raw Sample		3+ Eligible		3+ Eligible No Seniors	
Panel B: Household Member Characteristics	N=4935		N=4817		N=4252	
	Mean	SD	Mean	SD	Mean	SD
Female	.541	.498	.542	.498	.534	.499
Age	30.998	16.227	30.94	16.315	26.477	10.837
Literate	.502	.5	.506	.5	.564	.496
Currently in School	.311	.463	.317	.465	.356	.479
Self-Report on Labor	.182	.386	.176	.381	.171	.377
Relation to Respondent: Spouse	.206	.404	.191	.393	.189	.392
Relation to Respondent: Child	.392	.488	.4	.49	.447	.497
Relation to Respondent: Parent	.097	.296	.098	.298	.037	.189
Relation to Respondent: Sibling	.111	.315	.114	.318	.123	.328
Total Reported Job Activities	.745	.749	.734	.746	.744	.746
No Activities	.437	.496	.444	.497	.436	.496
One Activity	.384	.486	.381	.486	.386	.487
Two Activities	.176	.381	.173	.378	.175	.38
Three Activities	.003	.055	.002	.05	.003	.051
Household Farm Work among Working	.85	.357	.846	.361	.841	.366
Household Business among Working	.145	.352	.146	.353	.147	.354
Wage Work among Working	.247	.431	.249	.432	.253	.435
Labor Module Order Position	3.73	2.897	3.797	2.901	3.811	2.909

Notes: Column headers indicate sample. The analysis sample used in the analysis is reproduced for comparison in the final two columns. The first pair of columns characterize the full set of households and individuals in the raw data. The second two remove households with less than three individuals who participated in the labor module. This is because respondents are removed from the analysis of responses order and fatigue and the inclusion of household fixed effects requires a minimum of two additional people in a household who participated in the labor module in order to estimate. Finally, we remove seniors in the final two columns which both drops individuals above 60+ as well as households who, without seniors do not have sufficient members to estimate in the analysis.

Table A.2: Robustness to Sample Definition

| | Dependent Variable: Labor Activities Recorded | | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)
Order (1-14)	-0.017***	-0.017***	-0.019***	-0.016***	-0.016***	-0.018***
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.006)
N	4817	3879	3122	4252	4582	1938
Excluded Groups	Full	Self-Report	Students	Seniors: 60+	Incapacitated	All Restrictions
Mean Y	0.734	0.711	0.935	0.744	0.767	0.977
Percent Effect	-2.222	-2.216	-2.488	-2.179	-2.171	-2.400
Households	988	874	793	950	968	554
R2	0.523	0.534	0.433	0.578	0.561	0.507

Notes: All regressions include household, age group, and relation to respondent fixed effects as well as controls for gender, student, and being conferred with in the labor module. "Full" is the full set of individuals from the raw data. "Excluded groups" indicates sample exclusion criteria. Incapacitated are people revealed to be too old or unfit to work.

Table A.3: Work Types, Labor Supply, and Pay

	Type of Work			Pay Last Week			Hours Last Week		
	(1) HH Farm	(2) HH Bus	(3) Wage Work	(4) Any Pay	(5) IH(Pay)	(6) Log(Pay)	(7) Any Hours	(8) Hours	(9) Log(Hours)
Order (1-14)	-0.006**	-0.001	-0.005**	-0.006***	-0.035***	-0.020	-0.007**	-0.130	0.008
	(0.003)	(0.002)	(0.002)	(0.002)	(0.010)	(0.027)	(0.003)	(0.106)	(0.008)
N	4252	4252	4252	4252	4219	446	4252	4219	1112
Mean Y	0.474	0.083	0.143	0.162	0.698	3.771	0.341	10.600	3.159
Percent Effect	-1.222	-1.557	-3.775	-3.957	-	-	-1.920	-1.228	-
All	All	All	All	All	All	All	All	All	All
R2	0.571	0.389	0.388	0.438	0.442	0.734	0.489	0.478	0.753

Notes: This table looks at impacs of fatigue on additional labor related outcomes. Percent effects are not reported for outcomes converted with the inverse hyperbolic sine transformation (IH) or logs, as indicated in the column headers, as these percents can be read directly from the regression results.

Table A.4: Fatigue and Respondent Order by Household Size

	Order (1-14)			Percentile (0-1)		
	(1)	(2)	(3)	(4)	(5)	(6)
Order Position	-0.038**	-0.018**	-0.014***	-0.052	-0.101**	-0.133***
	(0.016)	(0.009)	(0.005)	(0.036)	(0.043)	(0.046)
N	1339	1383	1530	1339	1383	1530
Mean Y	0.786	0.747	0.705	0.786	0.747	0.705
Percent Effect	-5.100	-2.459	-1.910	-7.002	-13.628	-17.901
Family Size	2-4	5-7	8+	2-4	5-7	8+
Households	477	263	210	477	263	210
R2	0.649	0.561	0.537	0.648	0.561	0.537

Notes: Dependent Variable: Total Jobs Recorded. Columns (1)-(3) use the order number the household member was asked about in the labor module from 1-14. Columns (4)-(6) use the within household percentile position in place of the order number ranging from zero (first position) to one (last position).

Table A.5: Self Report Heterogeneity and Fatigue

Dependent Variable: Labor Activities Recorded		
	(1)	(2)
Order	-0.016***	-0.016***
	(0.004)	(0.005)
Order x Self Report		-0.001
		(0.013)
Self Report	0.073**	0.078
	(0.031)	(0.053)
N	4252	4252
Mean Y	0.744	0.744
Percent Effect	-2.183	-2.158
Household FEs	No	No
Family Size FEs	Yes	Yes
Enumerator FEs	No	No
Relation to Resp FEs	Yes	Yes
R2	0.578	0.578

Notes: Self Report indicates that this household member either reported their labor activities for themselves or were conferred with during the labor module.

Table A.6: Predictors of Self Reporting in Labor Module

	(1) Ghana - Analysis	(2) Ghana GLS 6	(3) Tanzania LSMS3	(4) Malawi IHS4
Female	0.031***	-0.123***	0.137***	0.151***
	(0.011)	(0.006)	(0.010)	(0.009)
Age 25-34	-0.096***	0.131***	0.045***	0.265***
	(0.018)	(0.010)	(0.015)	(0.017)
Age 35-59	-0.181***	0.373***	0.205***	0.376***
	(0.014)	(0.008)	(0.012)	(0.013)
N	5568	21321	8639	11900
Mean Y	0.213	0.432	0.702	0.400
R2	0.475	0.557	0.369	0.465

Notes: Dependent variable is an indicator for self-reporting during the labor module. Regressions include household fixed effects and in-school status. Data source indicated at the top of each column. The first column uses the data from our analysis.

Table A.7: Within Enumerator Heterogeneity

Dependent Variable: Labor Activities Recorded					
	(1)	(2)	(3)	(4)	(5)
Order	-0.016***	-0.016***	-0.018***	-0.016***	-0.018***
	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)
Order x Survey Not Enum's First of Day			0.003		
			(0.008)		
Survey Not Enum's First of Day		-0.017	-0.029		
		(0.023)	(0.035)		
Order x Second Half of Enum's Surveys					0.005
					(0.008)
Second Half of Enum's Surveys				-0.057***	-0.073**
				(0.022)	(0.034)
N	4252	4252	4252	4252	4252
Mean Y	0.744	0.744	0.744	0.744	0.744
Household FEs	No	No	No	No	No
Family Size FEs	Yes	Yes	Yes	Yes	Yes
Enumerator FEs	No	Yes	Yes	Yes	Yes
Relation to Resp FEs	Yes	Yes	Yes	Yes	Yes
R2	0.298	0.396	0.396	0.397	0.397

Notes: 55% of households were enumerator's first interview of the day. Columns (2) and (3) look at heterogeneity by whether the interview was after the enumerator's first of the day. Columns (4) and (5) look at heterogeneity by whether the surveys were after the midway point of the enumerator's total number of surveys through the full duration of the study. Without enumerator variation within households, these specifications replace household fixed effects used in other analyses with family size fixed effects.

Table A.8: Average Losses from Fatigue for Different Subgroups with Heterogeneity Models

Panel A: Gender Heterogeneity

	(1) Raw	(2) No Fatigue	(3) Fatigue	(4) % Diff
All	0.744	0.806	0.741	-8.043
	(0.746)	(0.566)	(0.574)	
Male	0.738	0.771	0.737	-4.305
	(0.726)	(0.561)	(0.566)	
Female	0.749	0.837	0.745	-11.047
	(0.763)	(0.568)	(0.581)	
Age 14-24	0.450	0.510	0.439	-13.843
	(0.649)	(0.473)	(0.477)	
Age 25-34	0.960	1.027	0.957	-6.771
	(0.737)	(0.476)	(0.479)	
Age 35-59	1.204	1.267	1.221	-3.641
	(0.652)	(0.421)	(0.429)	

Panel B: Age Group Heterogeneity

	(1) Raw	(2) No Fatigue	(3) Fatigue	(4) % Diff
All	0.744	0.805	0.741	-7.900
	(0.746)	(0.571)	(0.572)	
Male	0.738	0.797	0.736	-7.632
	(0.726)	(0.564)	(0.569)	
Female	0.749	0.811	0.745	-8.130
	(0.763)	(0.576)	(0.576)	
Age 14-24	0.450	0.495	0.443	-10.474
	(0.649)	(0.474)	(0.476)	
Age 25-34	0.960	1.081	0.955	-11.603
	(0.737)	(0.473)	(0.488)	
Age 35-59	1.204	1.236	1.213	-1.909
	(0.652)	(0.424)	(0.427)	

Notes: Panel A uses a model allowing for heterogeneity of the impacts of fatigue by gender following the specification in Column (2) of Table 5. Panel B uses a model allowing for heterogeneity of the impacts of fatigue by age group following the specification in Column (3) of Table 5.

The values in column (1) are means of the number of reported labor actvities in the raw data for individuals belonging to the group indicated at the start of each row. Column (2) removes predicted fatigue by setting $order_i = 0$ for all individuals then calculating the mean of these predicted "No Fatigue" values. Column (3) reimposes predicted fatigue if individuals had been asked about in the original, non-randomized, household listing order. Column (4) calculates the predicted proportion losses from fatigue.

ALL IFPRI DISCUSSION PAPERS

All discussion papers are available **here**

They can be downloaded free of charge

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
www.ifpri.org

IFPRI HEADQUARTERS

1201 Eye Street, NW
Washington, DC 20005 USA
Tel.: +1-202-862-5600
Fax: +1-202-862-5606
Email: ifpri@cgiar.org