The World Bank Economic Review, 00(0), 2024, 1–17
https://doi.org10.1093/wber/lhae040
Article

OXFORD

Private but Misunderstood? Evidence on Measuring Intimate Partner Violence via Self-Interviewing in Rural Liberia and Malawi

David Sungho Park, Shilpa Aggarwal, Dahyeon Jeong, Naresh Kumar, Jonathan Robinson, and Alan Spearot

Women may underreport intimate partner violence (IPV) in surveys. In an experiment in rural Liberia and Malawi, women were asked IPV questions via self-interviewing (SI) or face-to-face interviewing. Many respondents appear to misunderstand questions in SI, and significant effects of SI were observed on innocuous placebo questions. Because the prevalence of IPV is typically well below 50 percent, such measurement error will tend to bias IPV reporting upwards. Indeed, the results show that SI increases reported incidence of IPV, but it cannot be ruled out that these increases are spurious.

JEL classification: C93, O12, J16, I32

Keywords: intimate partner violence (IPV), measurement, ACASI, sensitive behaviors

1. Introduction

Intimate partner violence (IPV) is a pressing global public-health and policy problem, but measuring its true prevalence is challenging because factors such as social taboos, emotional pain, fear of retribution,

David Sungho Park is an assistant professor at the KDI School of Public Policy and Management, Sejong, Korea; his email address is park@kdis.ac.kr. Shilpa Aggarwal is an associate professor at the Indian School of Business, Hyderabad, India; her email address is shilpa_aggarwal@isb.edu. Dahyeon Jeong (corresponding author) is an economist in the DEC Development Impact (DIME) group at the World Bank, Washington, USA; his email address is dahyeonjeong@worldbank.org. Naresh Kumar is an economist at the World Bank, New Delhi, India; his email address is nkumar18@worldbank.org. Jonathan Robinson is a professor at the University of California–Santa Cruz, Santa Cruz, USA; his email address is jmrtwo@ucsc.edu. Alan Spearot is a professor at the University of California–Santa Cruz, Santa Cruz, USA; his email address is aspearot@ucsc.edu.
The research protocol for this study was approved by the IRBs of UCSC, the University of Liberia, and the Malawi National Committee on Research in the Social Sciences and Humanities (NCRSH). We thank USAID and IPA for funding. We are grateful to Jenny Aker for her collaboration. For organizing the data collection, we thank Joseph Davis, Arja Dayal, Wilson Dorleleay, Walker Higgins, Andreas Holzinger, Erik Jorgensen, Teresa Martens, Laura McCargo, and Camelia Vasilov at IPA Liberia, and Patrick Baxter, Emanuele Clemente, Calvin Mhango, Monica Shandal, Patrick Simbewe, and Asman Suleiman at IPA Malawi. We are extremely grateful to all the enumerators in both countries, though there are too many to list individually. We thank seminar participants at Baruch College, IFPRI, UCSC, IPA-GPRL Methods and Measurement Conference 2021, and KDI School-World Bank DIME Conference 2022 for helpful comments. We thank U.S. Agency for International Development and Innovations for Poverty Action for funding. A supplementary online appendix is available with this article at *The World Bank Economic Review* website. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of USAID, the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

© 2024 International Bank for Reconstruction and Development / The World Bank. Published by Oxford University Press

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

or feelings of shame or embarrassment cause women to hesitate in reporting IPV to friends or family, as well as to physicians or to law enforcement officials (WHO 2012; Garcia-Moreno et al. 2013). Spurred by the lack of systematic data on IPV and recognizing its epidemiological nature, organizations such as the WHO began to run large-scale, multi-country surveys to measure the prevalence of IPV in the 1990s (WHO 1996).[1] The latest estimates from these surveys reveal that more than a quarter of ever-partnered women globally have experienced physical or sexual IPV during their lifetime (Sardinha et al. 2022).

Many public-health professionals worry that the true rate of IPV may be higher, because women may understate their IPV experience even in surveys. It remains unclear whether this is the case. On the one hand, some of the stigmas that drive underreporting may be mitigated by the confidentiality afforded by a professionally conducted survey (as articulated in an informed consent form, for example), and by the fact that the surveyor is unlikely to be known by the survey respondent or her partner, or to have a reason to interact with the respondent again. The survey setting also differs critically from that in normal life because the survey *directly asks* about IPV, rather than leaving the onus of initiating the conversation to the woman herself.[2] On the other hand, some of the same stigmas may still apply; for example, the victim may feel ashamed about her situation, hesitate to confide in another individual, or be scared of being overheard (despite survey precautions to guard against this).

To address some of these concerns, an alternative approach is the use of confidential self-interviewing (SI). In this approach, women self-administer IPV questions privately, which ensures that their answers are shielded even from the enumerator.[3] This paper evaluates one such interviewing technique which is known as audio computer assisted self-interviewing (ACASI). In ACASI, respondents listen to pre-recorded questions via headphones and respond using a touchscreen device (in our setting, a tablet).[4] The enumerator has no interaction with the respondent during this part of the survey, other than to explain the module at the beginning, and to be available in case the respondent seeks clarification.[5]

The intent of ACASI is that it will destigmatize IPV reporting, which is expected to lead to an increase in reporting. However, there are two other factors which may also affect IPV reporting in ACASI (especially when benchmarked against a professionally administered survey with a trained enumerator). First, self-interviewing lacks any human element, and it is conceivable that respondents may be more inclined to report sensitive behaviors to a human interviewer, since the respondent may perceive the enumerator to be empathetic or have built a rapport with her over the course of the survey.[6] If this channel is present, ACASI will actually understate IPV.

1 For example, the WHO multi-country study on women's health and domestic violence was initiated in 1997 and the DHS program started collecting information on IPV in 1990 (the first IPV module was fielded as part of the standard DHS in Colombia).

2 In fact, the medical literature has identified one of the key measurement approaches for IPV to simply ask the person. The WHO also recommends direct questioning as the "gold standard" method of measuring IPV. See https://apps.who.int/iris/bitstream/handle/10665/85239/9789241564625_eng.pdf.

3 The answers are not fully anonymized, however, since researchers have access to this data later on, but instructions during this part of the survey, and the consent form, clearly indicate that this data will be kept securely, so any risk of data breach is remote. Further, the researchers would have no reason to interact directly with the respondent outside of this research setting.

4 Another reason to recommend ACASI is that it is virtually impossible for the interview to be overheard by anyone in close vicinity. However, this is not relevant in our experiment (or in any survey which uses best practice face-to-face interviewing), since the survey is always conducted privately.

5 In our survey protocol, the respondent could pause the module to ask questions, and the enumerator could help her to resume from where she left off.

6 Indeed, Ellsberg et al. (2001) compile anecdotes from debriefings of IPV survey enumerators in Nicaragua, recounting how they were moved or distressed by the respondents' IPV experiences in face-to-face interviews, and some even reported that respondents sought their counsel during or at the end of the IPV module.

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

A second factor, which is the focus of our paper, is that self-interviewing requires the respondent to understand the questions on her own, and to use the tablet, which may not be easy. This is an especially salient concern in the case of IPV, as the standard set of questions for measuring IPV has fairly complex and nuanced language, and therefore it may not be straightforward to grasp without the surveyor helping with interpretation. In almost every setting, misunderstanding will tend to cause IPV to be *over*reported. This is because IPV is measured through a module containing 20 questions which are later indexed into four main categories (controlling behavior, emotional violence, physical violence, and sexual violence). Typically, the mean of each of these individual yes/no questions is well under 0.5, so a woman who does not understand the module and randomly answers yes or no will tend to bias the level of reported IPV on any given question *upwards*. This bias will be amplified in the indexing (which is set equal to 1 if the respondent reported any form of IPV).

To shed light on these various channels, we conduct a measurement experiment within surveys collected as part of an evaluation of an unconditional cash transfer program in rural Liberia and Malawi (Aggarwal et al. 2024a, 2024b). Women were individually randomized into whether the IPV module was asked either via face-to-face interviewing (FTFI) or over self-interviewing (SI). Baseline IPV rates differ dramatically across the two samples: the proportion of women experiencing any type of IPV over the past year (measured in FTFI) is 20 percent in Malawi but 38 percent in Liberia; as such, we opt to present all results separately by country.[7]

This paper has four main findings. First, similar to several prior papers (discussed in more detail below), we check for respondents' understanding of the SI tool through five non-sensitive comprehension questions, for which the answer should universally be yes. These were administered to *all* respondents through SI, irrespective of the modality through which they were selected to subsequently complete the IPV module. The specific wording of these questions was developed over pre-testing and was meant to be understandable among the study populations. These questions were the following: (1) Are you a woman? (2) Do you live in [location where the survey is being conducted]? (3) In the past week, did you sleep, during day or night? (4) In the past year, did it rain in your village one time or more? and (5) Have you heard about the coronavirus? A sizeable fraction of women incorrectly answer no to these questions, with rates as high as 14–22 percent for the rain and sleep questions.[8] However, responses to even the most basic questions on gender and location are not unanimously affirmative: 2–5 percent of the women incorrectly answer the gender question and about 10 percent answer the location incorrectly. Overall, only 62–70 percent answer all five questions correctly, and only 84–88 percent answer the three simplest questions correctly (gender, rain, and knowledge of the coronavirus).[9]

Second, after the basic comprehension questions, we included a further set of innocuous "placebo" questions, randomized to be administered by either FTFI or SI. Because the survey method was randomized, we can estimate the placebo treatment effects of SI. The placebo includes four questions for which the answer could be yes or no: (1) Did you do any farm work in the past year? (2) Did you go to the market in the past week? (3) Will you, or anyone in your household, eat any [rice/maize] next week, one

7 The difference in rates between countries is also apparent in the most recent Demographic and Health Surveys (DHS), where 32 percent of women in our study region in Malawi and 57 percent in our study region of Liberia reported they had experienced any form of IPV in the past year (table S1.1).

8 The research team found such responses even during pre-testing, and so repeatedly refined the surveys, including iterating on the exact language of these questions; however, such reporting remained. While we have no definitive answer to why women answer as they do, one anecdote is that women with young children or who are nursing interpret "sleep" as being about getting restful sleep. A similar explanation for the rain question is that respondents could have interpreted it as getting "enough" rain.

9 In phone surveys that we ran with a random subset of these respondents in the immediate aftermath of the COVID-19 lockdowns (on topics unrelated to IPV), we found universal awareness about the virus and specifically, that it was called the coronavirus (Aggarwal et al. 2022).

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

time or more? and (4) Will you, or anyone in your household, eat any type of meat next week, one time or more? We find sizeable, statistically significant placebo effects for three of these four placebo questions in Malawi, and two of four in Liberia.

Third, we find that our screening questions were not effective in identifying women who understood the module: we find spurious placebo effects even among those who "passed" the screening questions (a result which is robust to various ways of defining "passing"). The fact that the screening questions fail to identify women who comprehend the module invalidates its use in our study. While this result is specific only to our study, and does not necessarily apply to other studies which use a screening approach, this finding does suggest that future studies might consider using placebo questions post-screening. In addition, it may be advisable for future researchers to explore more effective screening questions.

Fourth, we find that SI increases IPV reporting by a substantial amount, for all categories we measured (controlling behavior, and emotional, physical, and sexual IPV): on a given individual question, 7 percent of women in Malawi and 14 percent in Liberia report yes in FTFI, and SI increased this percentage by 5 percentage points in Malawi and 3 percentage points in Liberia. As an index, the effects were even larger, at least in Malawi, where the probability of emotional, physical, or sexual IPV increased by 5–10 percent points, on a base of 7–16 percent. In Liberia, the effects were more modest (and not always statistically indistinguishable from zero): 1–8 percentage points on a base of 7–34 percent, and significant only for sexual IPV. Naively interpreted, the increase in IPV we document would match the narrative that women are hesitant to report IPV, and that FTFI dramatically understates prevalence. However, we know from the comprehension questions that a significant portion of women do not seem to understand the ACASI module, and the effect sizes for our placebo results are similar to those for IPV. Our interpretation is that ACASI may not be appropriate for measuring IPV, at least for these populations, and researchers should be extremely cautious about using it. In this context, we cannot determine whether the increase we see is partially or entirely spurious.

This paper is related to a large but as yet inconclusive literature about the effects of ACASI on measuring sensitive behaviors. In regard to sensitive behaviors generally, not just limited to IPV, studies suggest that self-administration increases reporting (see Tourangeau and Yan (2007) for a review), though this is not universally the case.[10] While an increase in reporting of taboo or sensitive behaviors is in itself sometimes interpreted as indicative of more truthful reporting (under a prior that such behaviors are underreported), this is not necessarily the case. Researchers typically do not have an objective measure of the underlying behavior, and thus ultimately, it is not clear whether this increase is indicative of increased truthful reporting or other factors, like miscomprehension. A small set of studies on self-reported sexual behaviors (measured via FTFI as well as ACASI) has attempted "ground-truthing" by using bio-markers of sexual activity or of STDs (see Hewett et al. 2008; Minnis et al. 2009; Kelly et al. 2014). While all three find that reported sexual activity is higher when elicited via SI, biomarker-based validity was found to be mixed. Such ground-truthing is virtually impossible when measuring IPV experience, because (a) the standard way of measuring IPV is over the period of a year, and (b) objective biomarkers of IPV victimization do not exist.

In regard to the use of ACASI for IPV measurement specifically, there are several closely related papers which give conflicting evidence. Fincher et al. (2015) randomizes ACASI and FTFI among African American women in Women, Infants, and Children (WIC) clinics and find higher levels of IPV in FTFI. Cullen (2023) randomizes IPV measurement between FTFI, ACASI, and list randomization among Rwandan and Nigerian women, and finds that IPV rates measured by ACASI are similar to FTFI (though much higher

10 For example, Newman et al. (2002) find that ACASI increases reporting of stigmatized behaviors but decreases reporting of psychological distress among syringe exchange program participants.

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

in list randomization). Peterman et al. (2023) randomize ACASI and FTFI among women in Senegal and find higher reporting under ACASI.[11]

An important concern with ACASI is that people may not understand the method. To deal with this, several papers use an approach, similar to our paper, to screen people based on innocuous screening questions. Peterman et al. (2023) find that 11 percent of respondents do not pass basic screening questions, in the general vicinity of our findings (which range from 12 to 38 percent depending on how "passing" the screening is defined, as discussed above). More generally, a number of other studies exist which measure IPV using ACASI, and that employ screening questions (administering the module via FTFI in such cases). These include Dunkle et al. (2020), who allowed people to opt out if they did not understand or preferred not to use the device (24 percent of women); Falb et al. (2016), who tested ACASI with adolescent girls in the Democratic Republic of Congo (DRC) and in refugee camps along the Sudan–Ethiopia border, and report that self-reported *average* ACASI comprehension levels are 90 percent for the DRC and 75 percent for the Sudan–Ethiopia border, a level similar to our study; and Park and Kumar (2022), a concurrent study to ours in Monrovia, Liberia, find that 10 percent do not pass screening questions among women in urban areas with higher education. Even if a majority of women do understand questions, even a sizeable minority may drastically affect estimates, since the proportion of women experiencing a specific type of violence is typically low (for example, in our study, the proportion responding yes to physical or sexual IPV over FTFI was about 3–4 percent in Malawi and 4–9 percent in Liberia).

This paper is also related to a recent literature about survey methodologies more broadly that aim to preserve respondent confidentiality. The most common alternative involves indirect responses, such as list experiments or randomized response techniques.[12] There is no consensus on the efficacy of these methods.[13] In regard to IPV specifically, Agüero and Frisancho (2022) find no difference in IPV reporting between list randomization and FTFI among urban microfinance borrowers in Lima, Peru, while as mentioned previously, Cullen (2023) finds that list randomization dramatically increases IPV reporting.[14]

The rest of this paper is laid out as follows. Section 2 describes the experiment and data collection. Section 3 presents our main results. Section 4 discusses evidence on potential pathways and heterogeneity. Section 5 concludes.

2. Data and Experimental Design

2.1. Setting

The ACASI experiment we analyze was done as part of an endline survey for a cash transfer randomized controlled trial in Liberia and Malawi (the transfers were implemented by the NGO GiveDirectly, as part of a USAID-funded study (Aggarwal et al. 2024a)). The study takes place in Bong and Nimba counties in Liberia, and in Chiradzulu and Machinga districts in Malawi. The study includes 300 villages in each country, with half of the villages receiving cash transfers worth $500 on average. While we do not evaluate the transfers themselves in this study, one important detail is that villages were included in the study only if they fell below a population threshold, as measured in the most recent population census (in Malawi,

11 In closely related work, Punjabi et al. (2021) find that ACASI leads to greater reporting of gender-based violence among school children in Uganda. However, the context here is not that of violence by intimate partners.

12 In list experiments, yes/no questions about sensitive behaviors are included in a list with other innocuous binary-response questions, and subjects report the number of items for which the answer is yes, allowing the researcher to back out the population-level prevalence of a behavior without being able to identify whether a specific individual engaged in that behavior. The randomized response technique (RRT) bundles a question with a random event, such as the throw of a dice. For instance, respondents are instructed to report truthfully only if the die landed on a certain number.

13 See Höglinger and Jann (2018), Lensvelt-Mulders et al. (2005), and Blair, Coppock, and Moor (2020) for reviews.

14 Researchers have also tried other unconventional methods to measure IPV indirectly, such as asking female community leaders, but these efforts have not been very successful (Agüero et al. 2020).

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

the upper threshold was 100 households per village; in Liberia, it was 125). The reason for this is that transfers were given out universally in treatment villages, and so our partner NGO chose smaller villages to be able to preserve their liquidity.

In Liberia, we implemented the project in two waves: a smaller first wave (90 villages), which had its endline in late 2020, and a bigger second wave (210 villages), which had its endline in September–November 2021. Most of our ACASI protocols were developed, tested, and refined over the course of the Wave 1 endline. Therefore, this sample is excluded from our results, and our results for Liberia are restricted to Wave 2 only. In Malawi, all 300 villages were enrolled at once and the endline was in April–July 2021. Figure S1.1 presents the project time line.

In both countries, we attempted to enroll 10 households per village into data collection for program evaluation, though in some cases we were only able to enroll fewer households. Surveys were targeted at female heads of households, i.e., either the spouse or partner of the male head for dual-headed households or in rare cases, the solo head for female-headed households (as we show in table 1, 97 percent of the sample in both countries had a partner at the time of our study). Male heads of dual-headed households were interviewed only when the female was not present and could not be reached within a few days of the initial visit; when the male head was interviewed, the IPV module was not asked.

2.2. Questionnaire Design and ACASI Experiment

2.2.1. Measuring Intimate Partner Violence

To measure IPV, we employed WHO's standard Violence Against Women module.[15] The questionnaire includes 20 questions about experience with specific forms of violence, over a time period of 12 months prior to the survey. Following the literature, we group these questions into four categories: controlling behavior, emotional IPV, physical IPV, and sexual IPV. In conducting this module, we followed WHO's ethics protocol for IPV research, which includes hiring only female enumerators; training enumerators to safely conduct the interviews and to be prepared emotionally for the work; conducting all surveys privately; reiterating consent just before the IPV module; and providing all respondents with an information sheet that lists the services available for women experiencing IPV, including law enforcement and local hospitals (WHO 2016).

The IPV module was attempted to be administered to all women who had an intimate partner within the 12 months preceding the survey. A small percentage (less than 1 percent) of women refused the IPV questions entirely. In addition, for the ACASI experiment, we excluded all women who reported having vision or hearing impairment (since they could not take the ACASI module); these women were administered the module via FTFI and are not included in this paper. About 7 percent of our sample were excluded for this reason.[16] We also excluded the Liberia Wave 1 sample because most of the ACASI survey protocols were developed and refined throughout that period. With all these restrictions, the sample size for this paper is 2,998 women (1,737 in Malawi and 1,261 in Liberia).

IPV prevalence in each country can be estimated from the most recent Demographic and Health Survey (DHS), which was conducted in 2020 in Liberia and 2016 in Malawi. The DHS uses an identical module via FTFI, and shows notable differences between the two countries. In table S1.1, we show IPV prevalence for the country as whole, as well as within our study regions specifically. In the study area, the reported prevalence of any form of IPV is 32 percent in Malawi and 57 percent in Liberia. These figures are significantly higher than the IPV rates observed in our study sample, which are 20 percent for Malawi

15 The WHO's standard questionnaire for measuring IPV, which is widely used for measuring IPV, can be found at https://www.who.int/gender/violence/who_multicountry_study/Annex3-Annex4.pdf. Our survey module on IPV can be found in appendix S6.

16 In table S1.2 we show that the likelihood of these exclusions was balanced between treatment and control.

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

Table 1. Summary Statistics and Experimental Balance

	Malawi		Liberia	
	FTFI Mean [SD] (1)	SI – FTFI (2)	FTFI Mean [SD] (3)	SI – FTFI (4)
Panel A. Demographics				
=1 if currently married or has partner	0.97	−0.01	0.97	−0.00
		(0.01)		(0.01)
Age	37.97	−0.94	37.13	0.67
	[12.88]	(0.60)	[10.96]	(0.61)
Number of household members	5.03	−0.02	5.59	−0.09
	[1.78]	(0.09)	[2.27]	(0.13)
Panel B. Education and mobile phone ownership				
Years of education	5.22	0.01	2.44	−0.01
	[3.50]	(0.17)	[3.43]	(0.19)
=1 if able to write/read	0.66	−0.01	0.30	0.01
		(0.02)		(0.03)
=1 if has access to mobile phone	0.45	0.03	0.42	−0.01
		(0.02)		(0.03)
Panel C. Household wealth				
Food security index (*z*-score)	0.00	−0.09*	0.00	0.05
	[1.00]	(0.05)	[1.00]	(0.06)
Total expenditure (monthly)	26.03	−2.13*	65.71	−0.52
	[24.46]	(1.17)	[47.08]	(2.59)
Net value of durables, livestock, and financial asset	162.55	4.24	416.43	33.88
	[235.93]	(11.45)	[823.80]	(51.15)
Non-agricultural income (monthly)	10.27	0.96	7.84	0.85
	[16.73]	(0.81)	[20.52]	(1.13)
Panel D. Empowerment-related outcomes				
=1 if has her own income source	0.44	0.01	0.31	−0.03
		(0.02)		(0.03)
Age difference from spouse	2.94	−0.16	4.09	0.72
	[10.78]	(0.51)	[12.59]	(0.72)
Panel E. Treatment status in parent study				
=1 if in cash transfers villages	0.51	0.01	0.52	−0.02
		(0.02)		(0.03)
=1 if in market access treatment villages	0.36	−0.02	0.34	0.01
		(0.02)		(0.03)
Observations	1,737		1,261	

Source: Authors' analysis from authors' household survey.

Note: Sample is restricted to women with an intimate partner over the 12 months prior to the survey, and those who do not report any vision or hearing impairments. Columns 1 and 3 present the mean for the face-to-face interviewing (FTFI) group, and columns 2 and 4 show the difference between the self-interviewing (SI) and FTFI groups. Standard deviation is in square brackets in columns 1 and 3, and standard error in parentheses in columns 2 and 4. ***, **, and * represent significance at 1 percent, 5 percent, and 10 percent, respectively.

and 38 percent for Liberia; however, the stark difference between the two countries is present in both measurements.

2.2.2. ACASI Implementation

In ACASI, respondents listen to questions on headphones and answer questions privately on a tablet. In each country, audio readings of the questions were recorded by an enumerator who was chosen for having clear enunciation. The recorded audio files were uploaded to SurveyCTO, along with image files for choice options (i.e. yes/no/refuse to answer/don't know). As shown in fig. S1.2, the

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

resulting interface on the tablet has a speaker icon (which the respondent could touch to listen to the question) and four images (from which the respondent could choose her answer by touching the screen herself).

In the field, the enumerator explained how to take the module, and then demonstrated how to conduct the module by going through a handful of practice and demonstration questions with the respondent, and making sure that she could clearly hear the audio and accurately choose the option she intends to. When the respondent felt ready to take the actual module, the enumerator handed the tablet over to the respondent for her to take the module. In order to make sure that she had complete privacy while doing so, the enumerator kept sufficient distance to be unable to see the screen but remained in the same room or vicinity to be available to answer questions. When the respondent handed back the tablet, the screen was blank so that the responses were blinded to the enumerator.

2.2.3. Experimental Design

In each survey round for each country, half of the sample randomly received ACASI while the other half randomly received FTFI. The randomization was not made in advance but was done "in the field" via the electronic survey software itself (SurveyCTO). The randomization was not stratified by either of the underlying treatments (cash transfers or market access) in the cash transfer experiment (Aggarwal et al. 2024a, 2024b). However, table 1, panel E shows that the sample is balanced on these treatments.

However, before starting the IPV questions, *every* respondent was asked to take a set of "comprehension" questions via ACASI.[17] The answers to all of these questions are expected to be yes: (1) Are you a woman? (2) Do you live in [the county/district in which the survey is being conducted]? (3) In the past week, did you sleep, during day or night? (4) In the past year, did it rain in your village one time or more? and (5) Have you heard about the coronavirus?[18]

After answering the SI comprehension module, women began questions in their experimental group (either ACASI or FTFI). As discussed throughout the paper, this module included questions on IPV, but it also included placebo questions. These placebo questions were meant to be innocuous and free from any stigma or social desirability bias, and to be a further tool to calibrate the effects of ACASI. They included four questions: (1) Did you do any farm work in the past year? (2) Did you go to the market in the past week? (3) Will you, or anyone in your household, eat rice/maize in the next week, one time or more?[19] and (4) Will you, or anyone in your household, eat meat in the next week, one time or more? Though the wording of these questions may look cumbersome in English, the specific language was developed after pre-testing to best be understood by the study populations.

2.2.4. Other Subtreatments

To explore possible technical reasons for misunderstanding, we cross-cut multiple subtreatments. First, we randomized whether the yes or no option would appear at the top of the screen (fig. S1.3). This randomization was implemented in order to test whether respondents were more or less likely to pick the first option. Second, in order to examine possible learning effects in which respondents became more

17 These comprehension questions were added after piloting, when it became apparent that women were answering unexpectedly to innocuous placebo questions.

18 We also asked one question that would likely be answered no: Have you traveled outside the country in the past week? We do not use this in our main specifications, however, because some women could potentially travel across borders (especially in Nimba county in Liberia, which borders Guinea, and Machinga district in Malawi, which borders Mozambique), due to which "no" is not a perfect benchmark.

19 This question is about the staple food, which is maize in Malawi and rice in Liberia.

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

comfortable with the method with more experience, we randomized whether the placebo questions came before or after the IPV module.[20]

2.3. Summary Statistics and Balance Check

Table 1 shows summary statistics by country sample, as well as the difference between the SI and FTFI groups. Panel A shows household demographics. Because the sample is restricted to women with an intimate partner at any point during the past 12 months, the proportion of women who are currently partnered is very high (97 percent). The average respondent is about 37–38 years old and lives in a household with 5–6 members. Panel B shows education and mobile phone ownership. Average educational attainment is 5.2 years in Malawi and only 2.4 in Liberia. Sixty-six percent of women in Malawi are literate, compared to 30 percent in Liberia. Mobile phone ownership is similar in the two countries, ranging from 42 percent in Liberia to 45 percent in Malawi.

Panel C shows some indicators of household income and wealth: average total monthly household expenditures are $26 in Malawi and $66 in Liberia, or about $0.17–$0.39 in per capita daily expenditures. In Malawi, the average household reports about $160 worth of assets, compared to $420 in Liberia. Most of the households in the study villages are subsistence farmers, and the average monthly non-agricultural income measured in our surveys is $8–$10.

Panel D shows a few proximate indicators related to female empowerment. Forty-four percent of women in Malawi have their own income source, compared to 31 percent in Liberia. The age difference (in years) between husband and wife is 2.9 in Malawi and 4.1 in Liberia.

Of the 14 variables in this table, we find 2 for which the differences are significant at 10 percent in Malawi (food security and total expenditures, which are both lower in the ACASI group), and none in Liberia. While the randomization appears to show no cause for concern, we control for all variables reported in table 1 for the main analysis.[21]

3. Results

3.1. Comprehension Questions

We start by documenting responses to the five comprehension questions which were administered to all respondents via ACASI. Results, which are shown in table 2, suggest major cause for concern. Only 95–98 percent report being a woman, and 91–93 percent report living in their county/district of residence. Even more surprisingly, only 78–86 percent report that they slept in the past week and 83–85 percent report that it rained in the past year. While we do not have a good explanation for these results, an ex post explanation from some of our field staff was that some women interpreted the sleep question as "getting a good night's sleep," to which some women reported no. A parallel ex post explanation for the rain question is that women may have interpreted it as meaning whether it rained "enough." The reasons for misinterpretation of these questions notwithstanding, the bottom line is that even these simple questions were very likely misinterpreted, raising concerns about how well the more nuanced IPV questions would be understood.

Taking the questions together, we find that only 62 percent of respondents in Malawi and 70 percent in Liberia correctly answered all the questions. If we restrict to the three simplest questions (gender, location, and knowledge of the existence COVID), these numbers become 84 percent and 88 percent. Either way, these responses suggest that many women will not be able to use ACASI effectively, and that it will be difficult to estimate a population-level prevalence using ACASI. We are

20 While we ended up not including this in the analysis for this paper, we also randomized the order between the IPV questionnaire and another likely sensitive module, the nine-question Patient Health Questionnaire (PHQ-9), which measures depression. We control for this cross-randomization in relevant analysis shown in appendix S3.

21 Results without controls are qualitatively very similar.

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

Table 2. Self-Interviewing Comprehension Questions

	Mean (=1 if yes)	
	Malawi (1)	Liberia (2)
Questions for which answer should be yes:		
1. Are you a woman?	0.95	0.98
2. Do you live in [the county/district where the survey is being conducted]?	0.91	0.93
3. In the past week, did you sleep, during day or night?	0.78	0.86
4. In the past year, did it rain in your village one time or more?	0.83	0.85
5. Have you heard about coronavirus?	0.93	0.94
=1 if yes to all five questions	0.62	0.70
=1 if yes to questions 1, 2, and 5	0.84	0.88
Observations	1,737	1,261

Source: Authors' analysis from authors' household survey.
Note: These five questions were asked via self-interviewing (SI) of everyone included in the measurement experiment.

not able to robustly identify subpopulations where ACASI will work well, at least in our sample—we show correlations between "passing" the SI comprehension test (defined as answering yes to the three simpler questions) and respondent characteristics in table S1.3. In Malawi, we see that more educated and literate women are more likely to "pass" the comprehension test, but the R-squared is only 0.01–0.02. In Liberia, however, these correlations are insignificant. As discussed below, ultimately these correlations are too weak to be usefully used to identify a subgroup that understands the questions.

3.2. Placebo Effects

Next we examine effects of SI on placebo questions, and test whether we can use the comprehension questions to predict which women are more likely to answer correctly using SI. To do this, for each country sample, we run the following regression:

$$Y_i = \beta \mathrm{SI}_i + \boldsymbol{X}_i'\boldsymbol{\theta} + \varepsilon_i, \tag{1}$$

where $\boldsymbol{X}$ is a vector of covariates including all variables in table 1.[22]

Results are presented in table 3. In Malawi, treatment effects are significant for three of four outcomes; in Liberia, they are significant for two of four. These results strongly suggest that ACASI will tend to generate spurious treatment effects.[23]

As mentioned above, the key goal of the screening questions is to identify a group for whom placebo effects are non-existent. In particular, in table 4, we examine heterogeneity in treatment effects for those who "passed" the comprehension test described above (defined as answering correctly to the gender, location, and COVID questions). However, in both countries, we cannot reject equality of effect sizes for

22 Results without controls, which are essentially identical, are shown in appendix table S4.1, table S4.2, and table S4.3.

23 These results are conditional on answering yes or no; we drop those that report "don't know" or who refused to answer. In table S1.4, we examine whether the probability of responding yes or no is correlated with ACASI. We find a small increase in responding "don't know" or "refuse to answer" for the farm work and market visit questions in Malawi, but a decline for the question about eating maize. In Liberia, we only see a relatively small effect on eating rice. In appendix table S5.1, table S5.2, and table S5.3, we show unconditional results, imputing "don't know" and "refusal" responses with zeros instead of dropping them. These are virtually identical to the main results.

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

Table 3. Effect of Self-Interviewing on Placebo Questions

	Farm work (past year) (1)	Market visit (past week) (2)	Maize/rice (next week) (3)	Meat (next week) (4)
Panel A. Malawi				
SI	−0.02	0.09***	0.08***	0.11***
	(0.01)	(0.02)	(0.03)	(0.03)
FTFI mean	0.94	0.47	0.54	0.29
Observations	1,718	1,713	1,345	1,228
Panel B. Liberia				
SI	0.01	0.08***	−0.04***	−0.01
	(0.02)	(0.03)	(0.01)	(0.03)
FTFI mean	0.80	0.67	0.98	0.69
Observations	1,259	1,260	1,226	1,101

Source: Authors' analysis from authors' household survey.
Note: Regressions include individual controls (including all variables in table 1). SI: self-interviewing; FTFI: face-to-face interviewing. Robust standard errors in parentheses. ***, **, and * represent significance at 1 percent, 5 percent, and 10 percent, respectively.

Table 4. Effect of Self-Interviewing on Placebo Questions, Heterogeneity by "Passing" the Comprehension Section

	Farm work (past year) (1)	Market visit (past week) (2)	Maize/rice (next week) (3)	Meat (next week) (4)
Panel A. Malawi				
SI × comprehension pass (β)	−0.01	0.08***	0.09***	0.10***
	(0.01)	(0.03)	(0.03)	(0.03)
SI × non-pass (γ)	−0.05	0.11*	0.03	0.19***
	(0.04)	(0.06)	(0.07)	(0.06)
Comprehension pass	0.02	−0.02	−0.01	0.03
	(0.03)	(0.05)	(0.06)	(0.05)
FTFI × non-pass mean	0.92	0.45	0.54	0.23
p-value ($\beta = \gamma$)	0.341	0.687	0.409	0.199
Observations	1,718	1,713	1,345	1,228
Panel B. Liberia				
SI × comprehension pass (β)	0.02	0.07***	−0.03**	−0.01
	(0.02)	(0.03)	(0.01)	(0.03)
SI × non-pass (γ)	−0.01	0.17**	−0.08*	−0.02
	(0.07)	(0.07)	(0.04)	(0.08)
Comprehension pass	0.09	0.10*	0.02	0.03
	(0.06)	(0.06)	(0.03)	(0.06)
FTFI × non-pass mean	0.72	0.57	0.96	0.65
p-value ($\beta = \gamma$)	0.719	0.219	0.273	0.884
Observations	1,259	1,260	1,226	1,101

Source: Authors' analysis from authors' household survey.
Note: "Comprehension pass" is defined by selecting yes to questions 1, 2, and 5 in table 2. By this alternative definition, 84 percent in Malawi and 88 percent in Liberia are in the "comprehension pass" group. Regressions are at the respondent–question level. Regressions include individual controls (including all variables in table 1). SI: self-interviewing; FTFI: face-to-face interviewing. Robust standard errors in parentheses. ***, **, and * represent significance at 1 percent, 5 percent, and 10 percent, respectively.

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

Table 5. Effect of Self-Interviewing on IPV (Individual Questions)

	=1 if responded yes to *individual* question in the following category:				
	Controlling behavior (1)	Emotional IPV (2)	Physical IPV (3)	Sexual IPV (4)	All questions pooled (5)
Panel A. Malawi					
SI	0.09***	0.05***	0.01*	0.03***	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
FTFI mean	0.11	0.07	0.03	0.04	0.07
Number of individuals	1,715	1,711	1,712	1,709	1,716
Observations	11,887	6,802	10,181	5,095	33,965
Panel B. Liberia					
SI	0.05***	0.01	0.01	0.03***	0.03**
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
FTFI mean	0.20	0.19	0.09	0.04	0.14
Number of individuals	1,259	1,259	1,259	1,259	1,259
Observations	8,752	5,006	7,508	3,758	25,024

Source: Authors' analysis from authors' household survey.

Note: Regressions are at the respondent–*question* level (violence is not aggregated into indexes). See table 6 for results in which IPV questions are aggregated into indices. Regressions include question-level fixed effects and individual controls (including all variables in table 1). SI: self-interviewing; FTFI: face-to-face interviewing. Robust standard errors in parentheses. ***, **, and * represent significance at 1 percent, 5 percent, and 10 percent, respectively.

those that passed and those that did not, for any question.[24] Ultimately, we conclude that the screening questions were ineffective, at least in this setting and for the specific questions we used, in fully identifying those with high comprehension. Future work may usefully explore whether using a separate set of questions may be more effective.

3.3. Effect of ACASI on IPV Reporting

Next we show the ACASI effects on the main outcome of interest, IPV. We first estimate a regression at the *question* level:

$$\text{IPV}_{\text{iq}} = \beta \text{SI}_i + \boldsymbol{X}_i'\boldsymbol{\theta} + \psi_q + \varepsilon_{\text{iq}},$$

where IPV_{iq} is the binary indicator of whether individual i responded yes to question q, and ψ_q question-level fixed effects. All other notation is the same as equation (1). We report results separately for each category of IPV: controlling behavior, emotional IPV, sexual IPV, and physical IPV. In a second analysis, we estimate the same equation but for the IPV *index*, which is set equal to 1 if a respondent reported violence on *any* question in that category.

The question-level results are presented in table 5. For Malawi (panel A), all effects are statistically significant, and range between 1 (physical IPV) and 9 percentage points (controlling behavior). A specification that pools all question categories together (column 5) finds a 5 percentage-point increase in reporting, also significant. In Liberia (panel B), effects are slightly more modest, where only three of five coefficients are significant, and effect sizes range from 1 to 5 percentage points. However, these effect sizes are similar to those for the placebo effects, and therefore, we cannot rule out the possibility that they are spurious. The more modest effects in Liberia are generally consistent with the possibility that some women are randomly answering yes or no: such behavior will cause prevalence to be biased towards 50 percent and thus will cause more of a distortion when the true prevalence is further from 50 percent.

24 In the spirit of table S1.3, we explore heterogeneity checks by education, mobile phone access, literacy, and age (tables S1.5 and S1.6). However, we find no consistent pattern.

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

Table 6. Effect of Self-Interviewing on IPV Indices

	=1 if responded yes to *at least one* question in the following category:				
	Controlling behavior (1)	Emotional IPV (2)	Physical IPV (3)	Sexual IPV (4)	Any IPV (5)
Panel A. Malawi					
SI	0.18***	0.10***	0.05***	0.06***	0.13***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
FTFI mean	0.39	0.16	0.08	0.08	0.20
Observations	1,715	1,711	1,712	1,709	1,716
Panel B. Liberia					
SI	0.07***	0.04	0.01	0.08***	0.04
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
FTFI mean	0.57	0.34	0.23	0.07	0.38
Observations	1,259	1,259	1,259	1,259	1,259

Source: Authors' analysis from authors' household survey.

Note: IPV measures are indexed by category; index is set equal to 1 if the respondent answered yes to any question in the category. Regressions include individual controls (including all variables in table 1). SI: self-interviewing; FTFI: face-to-face interviewing. Robust standard errors in parentheses. ***, **, and * represent significance at 1 percent, 5 percent, and 10 percent, respectively.

These results are shown only for women that answer the questions: we drop women who say "don't know" or "refuse to answer." We examine the effect of dropping these women in table S1.7. The probability of responding "don't know" or "refuse to answer" is quite modest in FTFI (1–2 percent), but SI increases this probability by about 1 percentage point in Malawi and by a smaller amount in Liberia. In order to bound the possible effects of these non-responses, in appendix S5, we rerun our main results with these questions being counted as not having experienced IPV, and the results are very similar.

In table 6, we show results at the index level. The findings are qualitatively similar to those for individual questions, although results here differ dramatically by country. In Malawi, ACASI increases emotional IPV by 10 percentage points (base 16 percent), physical IPV by 5 percentage points (base 8 percent), and sexual IPV by 6 percentage points (base 7 percent). Across all forms of IPV (not including controlling behavior), ACASI increases prevalence by 13 percentage points, a 65 percent increase on the base of 20 percent. In Liberia, effects are positive but surprisingly much more modest: the index of any form of IPV increases by only 4 percentage points (on the much higher base of 38 percent), and is not statistically significant.

4. Investigation of Heterogeneity and Pathways

4.1. Debriefing: Did Technical Problems Impede Understanding?

A simple hypothesis for these results is that technical problems made it hard to understand or complete the ACASI module, and therefore a technically superior module may eliminate the purported miscomprehension. We believe that this is unlikely because we extensively pre-tested the modules before they were fully rolled out, especially after pilot results showed similar patterns to those reported here. We carefully tested that the audio instructions were well articulated and read at a reasonable speed, and refined the implementation over time. Nevertheless, technical difficulties could have remained.

To shed light on this possibility, we make use of debriefing information we collected. After the respondent handed back the tablet to the enumerator, the enumerator asked a handful of debriefing questions about whether the respondent had faced any technical or comprehension difficulties during the module, which we present in appendix S2. As shown in table S2.1, only 1–2 percent reported technical issues; most respondents could hear the module, and felt the recordings were slow enough to understand. In

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

table S2.2, we regress answers to these technical questions on passing the SI comprehension module. We find no correlation here, which is perhaps not surprising given the low level of technical difficulties. We find no evidence that simple technical problems were the explanation.

On the other hand, we show in table S2.3 that 8–12 percent reported comprehension difficulties with the module, in remembering which picture meant yes (a green check) and which meant no (a red cross), or in using the tablet. In table S2.4, we regress passing the comprehension module on these measures of self-reported comprehension. In both countries, we see that people who reported understanding the module were more likely to have demonstrated doing so in their answers to the module (though significantly so only for one measure). This is consistent with the idea that some people had trouble understanding how the module worked.

4.2. Subtreatments

We also randomized several subtreatments to evaluate technical components of the module (results are presented in appendix S3). First, to examine whether the location of the choice options on the screen affects reporting, we randomized the order of the yes and no options. This subtreatment was motivated by our suspicion that when in doubt, some women may have the tendency to simply choose the first option. We start by analyzing this for the placebo questions in table S3.1, and find evidence that respondents were more likely to choose yes when it appears at the top of the choice options in Malawi for two of four placebo questions, although not in Liberia. Surprisingly, however, in table S3.2 we find no evidence of the presence of such behavior in either country when it comes to the IPV questions. We have no good explanation for why this may be the case.

Second, in order to check for the possibility that respondents may get better at understanding the module with practice, we randomized the order between the non-sensitive placebo questions and the IPV questions. For half the sample (cross-randomized into FTFI and SI), the IPV questions came before the placebo ones, while for the other half this order was reversed. For the placebo questions (table S3.3), we find no effect of ordering, other than for the farm work question in Malawi. However, the effect goes contrary to the expected direction as the placebo effect of SI comes about when the placebos come later (i.e., practice does not help). That said, we do not wish to make much of this lone coefficient, as the placebo effect of SI is not significantly different between "placebos first" and "IPV first" in all other cases. For the IPV questions, we report coefficients in table S3.4, and find that IPV reporting increases for sexual IPV in Malawi if the placebos come first. Overall, for Malawi, the effect of SI on the probability of answering a question "yes" is about 4 percentage points if the IPV questions come first, but 6 percentage points if the placebos come first (p-value for difference = 0.217). We find no significant effect of the ordering in Liberia either. This finding is consistent with the possibility that survey fatigue causes measurement error to increase, though it is also possible that the increase in IPV is real and that women became more familiar with the module over the course of the survey. We leave a further investigation of this channel to future work.

5. Conclusion

In this paper, we test the efficacy of SI versus FTFI in eliciting truthful responses regarding IPV from female respondents in the context of a cash transfer experiment in rural Liberia and Malawi. Our results suggest that a substantial fraction of women do not understand SI, as evidenced by the fact that a sizeable minority answer innocuous screening questions incorrectly in SI, and that women respond differentially to placebo questions when administered by SI as opposed to FTFI. This lack of understanding will tend to *increase* IPV reporting, since the true rate of IPV on an individual question is typically much less than 50 percent. This spurious increase will tend to be more prevalent the further the true IPV prevalence is

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

from 50 percent; indeed, we find that the effect of SI is more pronounced in Malawi (where prevalence is lower) than in Liberia.

Our results raise concerns about the use of ACASI, at least in these settings. The most troubling implication of our results is that misunderstanding will affect IPV reporting in the same direction as will destigmatization. Because many researchers have a prior that IPV is underreported in FTFI, ACASI may appear to be a better method simply *because* it increases the rate of reporting. However, our paper suggests that researchers should be wary before coming to this conclusion.

Our paper suggests several directions for future research. First, there may be greater benefit from having well-trained, empathetic enumerators than from SI in the context of measuring IPV. For example, in a natural experiment in Serbia, respondents of a WHO-run IPV survey ended up getting randomly assigned to either a previously inexperienced but well-trained enumerator (training duration of 2.5 weeks) or to an experienced, professional enumerator, but with less than a day of IPV training.[25] While 21 percent of the women reported having experienced physical or sexual IPV to the untrained enumerators, 26 percent reported IPV to the trained ones (Jansen et al. 2004).

Second, if ACASI is used, researchers should implement screening and placebo questions as part of a standard protocol. That at least some women do not understand ACASI has been documented in a number of studies, not only including ours, and thus it is clear that these women should be removed from SI. What is less clear is whether these screening questions are effective in separating those who do not understand from those who do: in our study, even those who passed screening showed placebo effects, and thus the screening questions were not effective. However, the fact that we at least included the placebo questions allowed us to (ex post) examine the possibility of misreporting.

Nevertheless, our paper does not necessarily imply that IPV prevalence measured under FTFI is closer to the true (unknowable) prevalence. In particular, it is possible that IPV under FTFI is underreported due to stigma, and thus a subset of ACASI respondents randomly answering yes or no could yield an overall prevalence closer to the true rate. Of course, the point remains that such random guessing is a poor measurement method.

Overall, we conclude that non-conventional methods to collect data about stigmatized behaviors should be implemented with caution as they may open up unexpected channels of bias. In general, no single method may be a panacea when it comes to truthful elicitation from respondents, and therefore, it is advisable to accompany new methods with extensive testing and other ways of ground-truthing prior to widespread implementation.

Data Availability Statement

Data is available through Harvard Dataverse on this link: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/2GD3WE.

References

Aggarwal, S., J. Aker, D. Jeong, N. Kumar, D. S. Park, J. Robinson, and A. Spearot. 2024a. "The Dynamic Effects of Cash Transfers: Evidence from Rural Liberia and Malawi." Unpublished.

Aggarwal, S., D. Jeong, N. Kumar, D. S. Park, J. Robinson, and A. Spearot. 2022. " COVID-19 Market Disruptions and Food Security: Evidence from Households in Rural Liberia and Malawi." *PLOS ONE* 17(8): 1–19.

———. 2024b. "Shortening the Path to Productive Investment: Evidence from Input Fairs and Cash Transfers in Malawi." *Journal of Development Economics* 170: 103288.

25 This was done in an effort to speed up the fieldwork midway through surveying after the assassination of then Prime Minister Zoran Đinđić in March 2003.

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

Agüero, J. M., Ú. Aldana, E. Field, V. Frisancho, and J. Romero. 2020. "Is Community-Based Targeting Effective in Identifying Intimate Partner Violence?" *AEA Papers and Proceedings* 110: 605–9.

Agüero, J. M., and V. Frisancho. 2022. " Measuring Violence Against Women with Experimental Methods." *Economic Development and Cultural Change* 70(4): 1565–90.

Blair, G., A. Coppock, and M. Moor. 2020. " When to Worry About Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments." *American Political Science Review* 114(4): 1297–315.

Cullen, C. 2023. " Method Matters: The Underreporting of Intimate Partner Violence." *World Bank Economic Review* 37(1): 49–73.

Dunkle, K., E. Stern, S. Chatterji, and L. Heise. 2020. "Effective Prevention of Intimate Partner Violence Through Couples Training: A Randomised Controlled Trial of Indashyikirwa in Rwanda." *BMJ Global Health* 5(12): e002439.

Ellsberg, M., L. Heise, R. Pena, S. Agurto, and A. Winkvist. 2001. " Researching Domestic Violence Against Women: Methodological and Ethical Considerations." *Studies in Family Planning* 32(1): 1–16.

Falb, K., S Tanner, K. Asghar, S. Souidi, S. Mierzwa, A. Assazenew, T. Bakomere, *et al.* 2016. "Implementation of Audio-Computer Assisted Self-Interview (ACASI) Among Adolescent Girls in Humanitarian Settings: Feasibility, Acceptability, and Lessons Learned." *Conflict and Health* 10(32). https://conflictandhealth.biomedcentral.com/articles/10.1186/s13031-016-0098-1.

Fincher, D., K. VanderEnde, K. Colbert, D. Houry, L. S. Smith, and K. M. Yount. 2015. "Effect of Face-to-Face Interview Versus Computer-Assisted Self-Interview on Disclosure of Intimate Partner Violence Among African American Women in WIC Clinics." *Journal of Interpersonal Violence* 30(5): 818–38. 00026 Publisher: SAGE Publications.

Garcia-Moreno, C., C. Pallitto, K. Devries, H. Stockl, C. Watts, and N. Abrahams. 2013. *Global and Regional Estimates of Violence Against Women: Prevalence and Health Effects of Intimate Partner Violence and Non-partner Sexual Violence*. World Health Organization.

Hewett, P. C., B. S. Mensch, M. C. Ribeiro, H. E. Jones, S. A. Lippman, M. R. Montgomery, and J. H. van de Wijgert. 2008. " Using Sexually Transmitted Infection Biomarkers to Validate Reporting of Sexual Behavior Within a Randomized, Experimental Evaluation of Interviewing Methods." *American Journal of Epidemiology* 168(2): 202–11.

Höglinger, M., and B. Jann. 2018. "More Is Not Always Better: An Experimental Individual-Level Validation of the Randomized Response Technique and the Crosswise Model." *PLOS ONE* 13(8): e0201770.

Jansen, H., C. Watts, M. Ellsberg, L. Heise, and C. Garcia-Moreno. 2004. " Interviewer Training in the WHO Multicountry Study on Women's Health and Domestic Violence." *Violence Against Women* 10(7): 831–49.

Kelly, C. A., P. C. Hewett, B. S. Mensch, J. C. Rankin, S. L. Nsobya, S. Kalibala, and P. N. Kakande. 2014. " Using Biomarkers to Assess the Validity of Sexual Behavior Reporting Across Interview Modes Among Young Women in Kampala, Uganda." *Studies in Family Planning* 45(1): 43–58.

Lensvelt-Mulders, G. J. L. M., J. J. Hox, H. van der, G. M. Peter, and C. J. M. Maas. 2005. "Meta-Analysis of Randomized Response Research: Thirty-Five Years of Validation." *Sociological Methods & Research* 33(3): 319–48. 00448 Publisher: SAGE Publications.

Minnis, A. M., M. J. Steiner, M. F. Gallo, L. Warner, M. M. Hobbs, A. van der Straten, T. Chipato, *et al.*. 2009. " Biomarker Validation of Reports of Recent Sexual Activity: Results of a Randomized Controlled Study in Zimbabwe." *American Journal of Epidemiology* 170(7): 918–24.

Newman, J. C., D. C. Des Jarlais, C. F. Turner, J. Gribble, P. Cooley, and D. Paone. 2002. "The Differential Effects of Face-to-Face and Computer Interview Modes." *American Journal of Public Health* 92(2): 294–97.

Park, D. S., and N. Kumar. 2022. "Reducing Intimate Partner Violence: Evidence from a Multifaceted Female Empowerment Program in Urban Liberia." Unpublished.

Peterman, A., M. Dione, A. Le Port, J. Briaux, F. Lamesse, and M. Hidrobo. 2024. *Forthcoming, World Bank Economic Review* "Disclosure of Violence Against Women and Girls in Senegal."

Punjabi, M., J. Norman, L. Edwards, and P. Muyingo. 2021. "Using ACASI to Measure Gender-Based Violence in Ugandan Primary Schools." *RTI International*. Research Brief. RTI Press Publication No. RB-0025-2104.

Sardinha, L., M. Maheu-Giroux, H. Stöckl, S. R. Meyer, and C. García-Moreno. 2022. " Global, Regional, and National Prevalence Estimates of Physical or Sexual, or Both, Intimate Partner Violence Against Women in 2018." *Lancet* 399(10327): 803–13.

Tourangeau, R., and T. Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133(5): 859–83.

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024

World Health Organization. 1996. "Violence Against Women: WHO Consultation, Geneva, 5–7 February, 1996." https://iris.who.int/handle/10665/63277.

———. 2016. Ethical and Safety Recommendations for Intervention Research on Violence Against Women. *Building on lessons from the WHO publication Putting women first: ethical and safety recommendations for research on domestic violence against women.*

World Health Organization & Pan American Health Organization. 2012. Understanding and addressing violence against women : intimate partner violence. World Health Organization. https://iris.who.int/handle/10665/77432.

Downloaded from https://academic.oup.com/wber/advance-article/doi/10.1093/wber/lhae040/7876502 by Korea Development Institute user on 17 November 2024