

Method Matters

Underreporting of Intimate Partner Violence in Nigeria and Rwanda

Claire Cullen



WORLD BANK GROUP

Africa Region

Gender Innovation Lab

June 2020

Abstract

This paper analyzes the magnitude and predictors of misreporting on intimate partner and sexual violence in Nigeria and Rwanda. Respondents were randomly assigned to answer questions using one of three survey methods: an indirect method (list experiment) that gives respondents anonymity; a direct, self-administered method that increases privacy; and the standard, direct face-to-face method. In Rwanda, intimate partner violence rates increase by 100 percent, and in Nigeria, they increase by up to 39 percent

when measured using the list method, compared with direct methods. Misreporting was associated with indicators often targeted in women's empowerment programs, such as gender norms and female employment and education. These results suggest that standard survey methods may generate significant underestimates of the prevalence of intimate partner violence and biased correlations and treatment effect estimates.

This paper is a product of the Gender Innovation Lab, Africa Region. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The author may be contacted at claire.cullen@bsg.ox.ac.uk.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Method Matters:
Underreporting of Intimate Partner Violence in Nigeria and
Rwanda

Claire Cullen*

JEL: J12, J16, K42, O15

Keywords: Gender, Domestic violence, Measurement, Norms

*Cullen: World Bank Africa Gender Innovation Lab & Blavatnik School of Government, University of Oxford, claire.cullen@bsg.ox.ac.uk. Acknowledgements: I thank Julien Labonne, Stefan Dercon, Arthur Alik-Lagrange, Noam Angrist, Dara Kay Cohen, Cheryl Doss, Markus Goldstein, Summer Lindsey, Anandi Mani, Mũthoni Ngatia, Sreelakshmi Papineni, Rachael Pierotti, Simon Quinn, Diana Contreras Suarez, Alex Twahirwa, Julia Vaillant, and seminar participants at the World Bank Africa Region Gender Innovation Lab, Oxford CSAE, IFPRI, UC Berkeley, Yale EGEN, Australian Gender Economics Workshop, and SVRI for helpful comments and suggestions. I would also like to thank the respondents, and the staff at Laterite Ltd in Rwanda and TNS Global in Nigeria, and Victoria Isika, Jean Paul Ngabonziza, Maritetou Sanogo and Oluwatoyin Zakariya for their superb research assistance and data collection work. The impact evaluations through which this data was collected were conducted by the World Bank Africa Region Gender Innovation Lab and were funded by the World Bank's Nordic Trust Fund, the Swiss Development Cooperation and the Government of Rwanda, and in Nigeria, by USAID and the Umbrella Facility for Gender Equality, a World Bank Group multi-donor trust fund. The views presented in this paper are those of the author and do not represent those of the World Bank or its member countries. All remaining errors are my own. This study was approved by the Oxford University Central University Research Ethics Committee, HML IRB, and the Rwandan National Ethics Committee. This paper was previously circulated under the title "Truth be told: underreporting of intimate partner violence in Rwanda and Nigeria."

1 Introduction

It is challenging to accurately measure sensitive attitudes, behaviors, and experiences such as political preferences, prejudice, risky sex, and intimate partner violence (IPV). Often, researchers and policy makers must rely on self-reported survey data, yet such data may not be accurate. For sensitive topics in particular, there are many reasons why survey respondents may be reluctant to tell the truth, including a fear of being judged, endangered, or legally penalized (Blair, Coppock and Moor, 2018). If misreporting bias is significant and systematic, researchers and policy makers risk over or under-estimating the prevalence and welfare costs of sensitive issues, and drawing inaccurate conclusions about their causes and potential solutions.

In this paper, we compare self-reported sexual and intimate partner violence rates across three different survey methods. We assess whether survey methods that increase respondent privacy and anonymity affect IPV reporting compared to the status quo of direct face-to-face interviews.¹ In Rwanda and Nigeria, we randomly assigned women to answer questions about their experience of emotional, physical and sexual violence using one of three methods: an indirect method (list experiment) and two direct survey methods: face-to-face questions asked by an enumerator, or audio computer-assisted self-interview (ACASI) on an electronic tablet. Compared to the direct face-to-face method, the list experiment method provides additional confidentiality as respondents do not directly disclose their experience of violence. Instead, they give a response that allows researchers to calculate average prevalence across the whole sample, essentially making respondents' answers anonymous (Blair and Imai, 2012). The ACASI method affords more privacy than face-to-face methods as respondents answer questions themselves by listening to pre-recorded questions in a headset and then selecting their answer on a tablet.

Results suggest that estimated IPV prevalence varies depending on the survey method used to measure it. The most widely used method, face-to-face interview, resulted in the lowest IPV rates, followed by the more private ACASI method, and finally, the anonymous list method with the highest. In Rwanda, IPV rates were double, and in Nigeria, up to 39 percent greater when measured using the list method compared to the direct methods. There was no statistically significant difference between the list and direct methods for rates of non-partner sexual violence in Rwanda or sexual IPV in Nigeria. This lack of difference for sexual violence could be explained by a lack of statistical power in Rwanda, and a Nigerian context where sexual IPV was not significantly stigmatized.

We also find that misreporting is systematic and correlated with characteristics often theorized as causes of IPV and targeted in women's empowerment programs. The characteristics that correlate differently with IPV depending on the measurement method used include indicators of women's education and employment, community gender norms, vulnerability, and marital relationship quality. For example, in Rwanda, when IPV was measured using the direct methods, there was no relationship between IPV and conservative community gender norms. However, when IPV was measured using the list method,

¹In more than 60 developing countries, Demographic and Health Surveys (DHS) have included a standardized domestic violence module where enumerators directly ask women about their experience of a set of specific emotional, physical, and sexual abuse episodes by an intimate partner. Using the face-to-face method, the WHO estimates that approximately one in three women worldwide has experienced IPV in her lifetime (García-Moreno et al., 2013). IPV prevalence varies widely, though rates are generally highest in Africa (Devries et al., 2013; Global Burden of Disease, 2017). Researchers collecting IPV survey data take steps to minimize the risk of underreporting, for example by instituting ethics protocols about privacy during interviews, asking multiple questions to give respondents several opportunities to disclose violence, and using specially-trained female enumerators to conduct women's interviews (Heise and Hossain, 2017; WHO, 2016, 2001). However, it is still assumed that face-to-face data collection may yield underestimates given the sensitivity.

we find a large, positive relationship between IPV and conservative norms. This suggests that women living in the most conservative villages underreport IPV when asked directly and that the list method’s anonymity (but not ACASI’s privacy) makes them comfortable enough to report it. Also, in both Rwanda and Nigeria, the education-IPV relationship switches depending on how IPV is measured. There is suggestive evidence of a positive relationship between IPV and women’s education when IPV is measured directly, but a negative relationship when using the list method. These results suggest that the standard, direct methods currently used in most surveys risk generating significant underestimates of IPV prevalence. Further, social scientists risk estimating biased correlations and, if treatment is correlated with misreporting, estimating biased treatment effects and drawing inaccurate conclusions about IPV’s causes and the effectiveness of interventions to prevent it.²

This paper contributes to the growing social science literature studying alternative survey methods that may reduce misreporting (Karlan and Zinman, 2012; Blattman et al., 2016; Chuang et al., 2018; Blair, Coppock and Moor, 2018; Lépine et al., 2020), including on IPV measurement (Agüero and Frisancho, 2017; Peterman et al., 2017; Joseph et al., 2017; Bulte and Lensink, 2019; Agüero et al., 2020).

Recent studies on IPV measurement have compared respondents’ experience of violence across the face-to-face and list experiment methods. They found mixed results, with one study in Peru finding the same rates of IPV between face-to-face and list methods (Agüero and Frisancho, 2017) and others in Vietnam, India, and Nigeria finding a difference (Bulte and Lensink, 2019; Joseph et al., 2017; Banerjee, La Ferrara, and Orozco as cited in World Bank, 2017).³ However, in several of these studies, respondents were asked about their IPV experience using two methods within the same survey about similar, but slightly different, violence acts. This potentially limits comparability across methods. To avoid this issue, we conducted one of the first IPV measurement experiments that randomly assigns respondents to answer questions using only one method rather than asking respondents two overlapping questions using different methods. Also, this paper aims to contribute to the limited evidence on who misreports on sensitive questions and why, the implications of misreporting for bias, and the trade-offs between different survey methods that seek to address misreporting.

This is the first paper to compare the prevalence of sexual and intimate partner violence using three survey methods that offer respondents differing levels of privacy and anonymity, and to compare across two countries, and for women and men.⁴ It is also one of the first papers to study IPV misreporting in Africa and to demonstrate that misreporting is substantial in some contexts, varies by the type of violence, and is correlated with characteristics commonly targeted in women’s empowerment interventions.

The rest of this paper is organized as follows. Section 2 reviews the literature on the measurement of sensitive topics and IPV. Section 3 outlines the experimental design and describes the sample and data. Section 4 presents the results, and 5 concludes.

²A further motivation for assessing IPV measurement methods is that the status quo of asking direct face-to-face questions places a high emotional burden on both respondents and enumerators. WHO guidelines thus recommend particular protocols for IPV research, for example, ensuring respondents have access to referral services (WHO, 2016). If an alternative measurement method was identified that was less biased, logistically easier and cheaper, and more ethical than the status quo, then it would likely be adopted by more researchers and contribute to new research on IPV. This would improve our understanding of the causes of IPV and may broaden the range of interventions that could be used to reduce it (Peterman et al., 2017).

³A related study by Agüero et al. (2020) in rural Peru compares women’s direct self-reported experience of IPV with a listing of IPV survivors by female community leaders. They found that community leaders significantly underreport IPV rates compared to women’s own self-reports.

⁴In Rwanda, we asked men about their perpetration of emotional violence using one of the three methods. The ACASI method was only tested in Rwanda.

2 Survey methods to measure sensitive topics

There is robust evidence that people underreport the truth about sensitive topics when they are directly asked in a survey. Such research compares self-reported survey answers to observed or independently verifiable outcomes such as tax records, receipt of welfare benefits, drug prescriptions for mental illness, and voting in an abortion law referendum (Bharadwaj, Pai and Suziedelyte, 2017; Meyer and Mittag, 2019; Gottschalk and Huynh, 2010; Murray-Close and Heggeness, 2018; Rosenfeld, Imai and Shapiro, 2016).

However, reliable, unbiased observational data are rare for many sensitive topics, including IPV.⁵ This is particularly true in developing countries, where there is limited available administrative data from the medical or legal services that IPV survivors may use. Further, where these services are available, the people able or choosing to access them are likely not representative of the broader population of those experiencing IPV.⁶

As a result of these administrative data limitations and concerns about underreporting on face-to-face surveys, researchers are considering alternative survey methods to minimize misreporting bias. Two popular alternative methods, ACASI and the list experiment, address slightly different motivations for misreporting. The ACASI method is primarily designed to overcome misreporting driven by shame when responding face-to-face to an enumerator, while the list experiment addresses misreporting driven by a preference for anonymity or fear of repercussions if one’s answer becomes known, as well as shame. Table 1 shows these methods by their approach to questioning and administration.

Table 1: Method comparisons by questioning method and administration

	Question & answer through enumerator	Question & answer through tablet
Direct questioning method	Face-to-face (status quo)	ACASI
Indirect questioning method	List experiment	

2.1 More Privacy: Audio Computer-Assisted Self-Interview

ACASI is a method of data collection where respondents listen to pre-recorded questions through headphones and respond by selecting their answers (for example, the corresponding colored square) on a touchscreen or keypad.⁷ After demonstrating how to use the technology, enumerators stand a small distance from the respondent and ensure that privacy is maintained.

ACASI eliminates the embarrassment that respondents may feel about disclosing a sensitive answer face-to-face with an enumerator, potentially minimizing underreporting (Falb et al., 2016; Jarlais et al., 1999; Jaya, Hindin and Ahmed, 2008; Langhaug, Sherr and Cowan, 2010). If shame is a key driver of underreporting, then reporting under ACASI should be greater than face-to-face. ACASI may also reduce enumerator-associated variance by providing a more uniform approach to data collection.

⁵Tourangeau and Yan (2007) define sensitive questions as those a respondent deems intrusive; where they fear the threat of disclosure and the possible consequences of giving a truthful response; or where there are social norms that imply there are socially desirable and undesirable responses.

⁶Palermo, Bleck and Peterman (2014) analyzed data from 24 developing countries DHSs and found that only 7 percent of women who report experiencing IPV made a report to a formal source that would then be captured in administrative data.

⁷Figure A1 in the appendix shows ACASI being used.

Several experimental studies have compared the prevalence of sensitive health behaviors between ACASI and face-to-face survey methods. Most found that respondents reported sensitive health behaviors and experiences, including IPV, at higher rates under ACASI than the face-to-face method, though the evidence is mixed (Fay et al., 1998; Jaya, Hindin and Ahmed, 2008; Langhaug, Sherr and Cowan, 2010; Mensch et al., 2008; Phillips et al., 2010; Fincher et al., 2015). In one of the few method comparisons with an objective indicator, in Brazil, Mensch et al. (2008) compared sexually transmitted infection bio-markers with face-to-face and ACASI self-reports of risky sexual behavior. They found more underreporting when respondents were interviewed face-to-face than with ACASI.

2.2 More Confidentiality and Anonymity: List Experiment

Researchers have also used the list experiment to improve data accuracy when measuring sensitive behaviors and attitudes.⁸ The standard list experiment works as follows: respondents are randomly assigned to one of two groups to answer the list question. In the ‘non-sensitive’ or ‘control’ group, enumerators read respondents a set of three non-sensitive statements and ask the respondent to tell them how many but not which statements are true for the respondent, with possible answers between zero and three. Meanwhile, the ‘sensitive’ or ‘treatment’ group is read the same list of statements, with the addition of a fourth sensitive statement. They are given the same instructions, to say how many, between zero and four statements, are true. Given the sensitive and non-sensitive list groups are randomly assigned, the two groups are, in expectation, equivalent and the only difference is the inclusion of the sensitive statement. The prevalence of the sensitive item across the whole sample is thus the difference in the mean number of items reported by the two groups.

Unlike with the ACASI and face-to-face methods, anonymity is assured because the lists are designed in such a way that very few people in the list sensitive group have all four statements as either true or not true.⁹ This way, a respondent is not revealing their true answer to the sensitive item- to the enumerator or anyone with data access. This could mitigate underreporting by respondents concerned about consequences if someone were to find out their answers, or those who doubt the confidentiality of their data, or who find it easier to truthfully admit sensitive answers when asked indirectly.

Ex-ante, it is not clear whether shame or fear would be the biggest driver of misreporting on IPV survey questions. Assuming that responding in the affirmative is stigmatized, if the shame from answering a question face-to-face dominates other motivations for underreporting, then reporting is likely to be greatest under ACASI. If concerns about being identified or fear of potential repercussions dominate, then reporting would likely be greatest under the list experiment.

There are potential drawbacks to the list experiment. One challenge is that, by design, it cannot be used to identify individuals who respond affirmatively to the sensitive question of interest, potentially limiting its use where individuals need to be referred to support services, for example. Also, with list experiment answers ranging from zero to four, it typically has greater variance than direct questioning

⁸For example, Miller (1984) developed the list method to study illicit drug use. Others have used it to study sensitive topics including prejudice in the United States (Kuklinski, Cobb and Gilens, 1997), political preferences in Lebanon (Corstange, 2009), support for militant groups in Afghanistan (Blair, Imai and Lyall, 2014), risky sexual behavior (Chuang et al., 2018; Jamison, Karlan and Raffler, 2013; Lépine et al., 2020), support for female genital cutting (De Cao and Lutz, 2018), vote buying (Gonzalez-Ocantos et al., 2012), and use of microfinance loans (Karlan and Zinman, 2012).

⁹This can be done through careful piloting and selecting a set of statements that are all plausible but where the co-occurrence of all being true or untrue for any one person is unlikely. See Glynn (2013) and Corstange (2009) for a discussion and suggestions.

methods (but also likely lower misreporting bias because of respondent anonymity). Thus list experiments may reduce bias at the cost of efficiency (Blair and Imai, 2012).¹⁰

Recently, researchers have begun using the list method to measure IPV and other forms of gender-based violence. In several studies, IPV rates are compared between the face-to-face and list methods. The study most closely related to this paper, conducted in Lima, Peru, found no difference in overall IPV prevalence by survey method (Agüero and Frisanco, 2017); studies in other contexts did find a difference. In Sri Lanka, Traummüller, Kijewski and Freitag (2019) found the self-reported experience of wartime sexual violence (not IPV) to be higher under the list method than face-to-face, and in India, Joseph et al. (2017) found higher rates of IPV under the list method than face-to-face.¹¹

In these studies, respondents were either a) randomly assigned to answer questions about the same experience of IPV using the direct face-to-face method or the list method (Agüero and Frisanco, 2017; World Bank, 2017), or b) all respondents were asked about their IPV experience using both methods within the same survey- either with similar, but slightly different, question wording (Bulte and Lensink, 2019) or the same question wording (Joseph et al., 2017).¹² One issue with approach b) is that being asked the same question in two different ways may bias responses to the second question. We take approach a), asking individuals each question only once to minimize this risk, and we also test three methods with varying degrees of privacy and anonymity to provide additional insights into the reasons why people misreport.

2.3 Misreporting and truth

Tourangeau and Yan (2007) argue that for sensitive questions where responding in the affirmative is taboo, the survey method that finds the highest prevalence is closest to truth. In this paper, particularly given the stigma surrounding the issue of IPV in the study contexts, we assume that in most cases, the method with the highest prevalence is closest to truth and that most misreporting is underreporting. In rare cases, people may have incentives to overreport, for example, if a respondent believed she would have a greater chance of accessing a program if she reports experiencing IPV. However, it seems likely that for most sensitive, stigmatized topics such as the perpetration and experience of IPV, people are more likely to under than overreport.

Without objective, non-self-reported data, we cannot, ultimately, confirm which method is closest to truth. However, the two alternative survey methods tested in this paper are designed to address several reasons why people might underreport when interviewed with the face-to-face method- shame and fear.

¹⁰Blair, Coppock and Moor (2018) illustrate this trade-off in their review of the literature comparing the list method versus direct questioning experiments in social psychology and political science.

¹¹Aside from the Nigeria study (Banerjee, La Ferrara and Orozco, 2019; World Bank, 2017), one other study in Africa has used the list method to measure IPV. Peterman et al. (2017) measured the prevalence of physical violence using the list method in Zambia. They found rates of IPV similar to those reported in the Zambian DHS, which used the face-to-face method, though the samples are not directly comparable.

¹²For example, in Bulte and Lensink's (2019) Vietnam study, the sensitive list method group responded to a list that included the sensitive statement "I am regularly hit by my spouse," then all respondents later in the survey were also asked "how often did your husband push, slap, beat or hit you during the last 6 months" with response options never, rarely, sometimes, often, very often, or refuse. With this design, the questions may not be directly comparable given potential differences in interpretation and biased responses to the second question. Similarly, Joseph et al. (2017) first asked all respondents using the list method, with the sensitive statement being: "At least one woman member of my household has faced physical aggression from her husband anytime during her life," followed by all respondents directly being asked: "Has at least one woman member of your household faced physical aggression from her husband anytime during her life?" Further, this latter study departs from IPV research guidance, which advises asking about specific, unambiguous episodes of violence that minimize the risk of subjective interpretation (Heise and Hossain, 2017). The survey was also administered at the household level implying that respondents could be men or women, potentially including IPV perpetrators.

Further, the alternative methods are unlikely to introduce new reasons for people to report IPV that they have not experienced. This is particularly true for the list experiment, where answers are anonymous by design, hence individuals have no incentive to overreport given they cannot be followed-up with to provide aid or program services. Therefore, it seems likely that the method with the highest reporting indicates the method that best addresses respondents’ reasons for misreporting and is closest to truth.

This assumption is supported by empirical evidence from the only study we identified that compares an observed sensitive behavior with self-reported behavior measured using direct and indirect methods including the list experiment. Rosenfeld, Imai and Shapiro (2016) found that when they were asked directly, people underreported their past voting behavior on a sensitive abortion referendum in Mississippi. However, rates were higher and closer to truth when they were asked using the list experiment. No method yielded rates higher than the true ‘socially undesirable’ outcome of voting against tightened abortion restrictions. This study suggests that indirect methods such as the list experiment, which mask a respondent’s true answer, can reduce misreporting bias.¹³

Overall, evidence on the measurement of sensitive topics suggests that misreporting appears to be highly issue and context dependent (Blair, Coppock and Moor, 2018; Langhaug, Sherr and Cowan, 2010; Tourangeau and Yan, 2007), presumably affected by factors such as the interview setting, cultural norms, security and safety, and confidence in data privacy. Therefore, this paper primarily relates to misreporting about intimate partner and sexual violence, and in contexts similar to those studied here.

3 Experimental design

3.1 Context, data & design

Rwanda and Nigeria have high rates of IPV compared to the global average. According to data gathered using the direct face-to-face method, 45 percent of Rwandan women and 31 percent of Nigerian women report having experienced physical or sexual IPV in the past 12 months (Global Burden of Disease, 2017). This compares to 13 percent of women in the United States and United Kingdom, 42 percent in Kenya, 39 percent in India, and 31 percent in South Africa. In terms of IPV predictors, poorer, less educated women report IPV at higher rates in Rwanda, while richer, more educated women report more IPV in Nigeria (DHS, 2016, 2014).

Both experiments described below were conducted in accordance with WHO guidelines on the safe and ethical collection of data on violence against women (WHO, 2001, 2016).

3.1.1 Rwanda

In Rwanda, the measurement experiment was conducted on a sample of 2,728 heterosexual couples who were participating in the baseline survey of an IPV prevention study. To be eligible for the intervention, couples had to be aged 18 years or over and married or cohabiting for at least 12 months. The study took place in Eastern Province, in eight sectors selected by the Government of Rwanda for high rates of IPV and a high concentration of Village Savings and Loan Associations (VSLAs) (which the intervention was

¹³It also demonstrates that experiments that compare reporting rates across methods can put a sign and bounds on the likely misreporting bias of standard survey methods.

to be delivered through).¹⁴ The sample comprised mostly poor rural households, with women’s mean age of 37 and a median of 4 years of schooling.

For the measurement experiment, individual respondents were randomly assigned to answer two experimental questions using one of the three survey methods: face-to-face, ACASI, or list experiment. The sample was equally split across the three groups.

The two experimental questions were asked approximately one hour into the survey, immediately after ACASI-administered questions from the modified Conflict Tactics Scale about all respondents’ experiences of IPV in the past 12 months (for women), and the likelihood of becoming violent in a set of hypothetical circumstances (for men).¹⁵ Respondents were then asked the experimental questions according to the method they were assigned. The questions were:

Women:

1. In the past 12 months, has your husband pushed or thrown you against something like a wall or piece of furniture?
2. Has someone other than your current husband ever forced you to have sex with them when you did not want to?

Men:

1. In the past 12 months, have you tried to limit your wife’s contact with her family?
2. In the past 12 months, have you threatened to hurt your wife or someone close to her?

The surveys were administered by enumerators the same sex as the respondent, in private settings where interviews could not be overheard. Most surveys were conducted in empty school classrooms, with one enumerator and one respondent per classroom.¹⁶ The empty classroom interview sites afforded a more private setting than is achieved in many DHS and other household survey settings. Therefore, if a lack of privacy causes women to underreport, then the prevalence differences seen in this more private study setting between face-to-face and the two more private methods are likely to be smaller than the differences that might be observed in standard interview settings.

¹⁴Sectors are administrative units two levels below province, and 2 levels above village. Within the selected sectors, existing VSLAs were identified in 128 villages. The researchers ruled out 30 villages with fewer than three VSLAs each, and villages that were very close. This was done to minimize the risk of cross-village contamination for the separate randomized controlled trial, detailed in the RCT pre-analysis plan (Alik-Lagrange et al., 2017). Within the 98 study villages, the final sample included 2,042 intervention-eligible couples where at least one partner was a member of a VSLA, and 686 couples who were community members selected at random from a village household roster and who were ineligible for the intervention. VSLA couples’ treatment status was assigned by the research team only after baseline surveys were completed. Assignment was based on their ranking in a public lottery process that they applied to. The intervention, called ‘Indashyikirwa 2’, is a weekly couples group discussion that encourages couples to question traditional gender norms and build communication skills. Data were collected from December 2017 to February 2018.

¹⁵When all respondents reached the main section on IPV (which was to be asked through ACASI), they were shown how to use the ACASI module on the tablet and then instructed through practice questions. After the enumerator was satisfied that the respondent could use ACASI, respondents were left to answer the module while the enumerator stood approximately 5 meters away. Over 95% of female respondents and 98% of male respondents got the first ACASI practice question correct. After being assisted again, 99% of male and female respondents were able to answer the ACASI module in a way that maintained their privacy but with some support, for example, if they could not touch the tablet screens, they listened in the headphones and told the enumerator which button to select for them. The IPV ACASI questions asked of all respondents will be used in the randomized controlled trial. For men, the two experimental questions were the only IPV perpetration questions asked in the survey. The survey themes and order of questions are listed in Table A1 and Table A2.

¹⁶When classrooms were not available, surveys were conducted in empty offices, churches, or outside but strict privacy protocols were adhered to, including the requirement that interviews pause if someone else was within earshot or eyesight.

3.1.2 Nigeria

In Nigeria, the measurement experiment was conducted during the endline survey wave of a cluster randomized control trial studying the effects of the Feed the Future Nigeria Livelihoods Project, a multi-component cash transfer, nutrition, and agriculture program.¹⁷ The survey took place in Northwest Nigeria on the border region with Niger. The sample consisted of ultra-poor households, with approximately 30 percent in polygamous marriages, 84 percent identifying as Muslim, and 15 percent of women self-reporting as literate.

The eligibility criteria for the measurement experiment were that the respondent was female, over 17 years of age, and married and living in the same household as a male. In total, 2,817 women met these criteria and were randomly assigned to answer three experimental questions about their experience of IPV using either the face-to-face or list method. Assignment to the method groups was stratified on the same geographic and poverty-related strata used to assign respondents to treatment arms for the randomized control trial.

The three IPV questions used for the measurement experiment were either directly from or combined several violence questions from the DHS.¹⁸ The experimental questions were:

1. In the past 12 months, has your husband said or done something to humiliate you in front of others?
2. In the past 12 months, has your husband slapped you, pushed you, shaken you, or thrown something at you?
3. In the past 12 months, has your husband physically forced you to have sexual intercourse with him when you did not want to?

The respondents were interviewed one-on-one by a female enumerator in a private setting, usually in the household compound. Enumerators were instructed to pause if anyone came within earshot of the interview, not proceeding until privacy was assured.¹⁹

3.2 The measurement experiment

We assess misreporting bias by comparing rates of IPV and sexual violence across three different survey methods: face-to-face, list experiment, and ACASI (with the latter only used in Rwanda). Given that assignment to the survey method was random, we assume that the only difference in reported violence between the groups is due to the method used.

3.2.1 Estimation: Calculating prevalence under each survey method

Prevalence under the ACASI and face-to-face methods, where the sensitive item is asked directly, is simply the mean proportion of respondents who said they had experienced violence.

¹⁷The study evaluated the impacts of a package of livelihoods interventions, unconditional cash transfers, and a mentoring program, and was implemented by Catholic Relief Services (Bastian, Goldstein and Papineni, 2017). The cash transfers and most livelihoods programming ended in 2017, approximately 12 months before these data were collected.

¹⁸Female respondents were first asked about their experience of IPV using the list method (asked either the sensitive or non-sensitive list, depending on her randomly assigned group), followed by direct face-to-face IPV questions. The direct questions included the three measurement experiment questions (asked only to those in the face-to-face assigned group), and five face-to-face questions asked of all respondents (to be used as outcome measures for the main randomized trial).

¹⁹Occasionally this was difficult to achieve in busy household compounds, for example, when household members came to the house to collect items, causing enumerators to pause interviews. Data were collected from May-July 2018.

Under the list method, prevalence of the sensitive statement is calculated by taking the difference in the mean number of statements reported as true by the two groups (Blair and Imai, 2012). There are two approaches to administering the list to the non-sensitive group. Under the original approach, which we used in Nigeria, respondents are read the same list, bar the fourth sensitive statement, and asked to report how many of the three statements apply to them. Under the alternative design proposed by Corstange (2009), which we used in Rwanda, the non-sensitive group is read out each non-sensitive statement and asked directly whether or not each was true for them, yes or no. Individuals’ responses were then summed to create an equivalent count of three non-sensitive statements per respondent, from which prevalence can be estimated in the same way, by calculating the difference in means.²⁰

To help respondents keep track of their answers, they were advised to put their fist behind their back or under their hijab out of sight of the enumerator and to straighten a finger each time a true statement was read out. Once all statements had been read out, respondents showed the enumerator their hand to count the number of straightened fingers.²¹

Prevalence under the list method was calculated using a multivariate regression model, following the approach of Agüero and Frisancho (2017). The OLS regression estimated was as follows,

$$Y_i = \alpha + \rho T_i + \epsilon_i$$

where ρ is the coefficient of interest, the difference in the mean number of items reported across the two groups, or prevalence. Y_i is the number of statements on the list that are true for individual i , α is the mean number of items for the non-sensitive group, and T_i is assignment to the list method treatment group.²² The list non-sensitive group comprised the share of the sample that answered the experimental questions using ACASI or face-to-face. Table 2 shows the method assignment and group size.²³

Table 2: Method assignment group size

Answer type	List experiment	Face-to-face	ACASI
	Indirect	Direct	Direct
Rwanda			
<i>N</i> : women	873	918	937
<i>N</i> : men	877	909	942
Nigeria			
<i>N</i> : women	1,405	1,412	

Overall, both samples were well balanced on key observable characteristics. Results in Tables A5 to

²⁰We used Corstange’s approach in Rwanda because of early concerns about statistical power for conducting multivariate analysis, and this alternative approach allows an associated estimator that can be more efficient than OLS. Given the different construction of the lists, this alternative design potentially poses the risk that respondents do not answer in the same way across the sensitive and non-sensitive lists.

²¹In Rwanda, respondents also put their hands behind their backs and transferred a stone from one hand to the other each time a true statement was read out.

²²Power calculations indicated that the sample in Nigeria was large enough to detect a difference of approximately 6 percentage points between the groups, and in Rwanda, approximately 5 percentage points, with a significance level of 0.05, power of 0.8, and base prevalence of 20 percent.

²³Table A3 shows the wording of the questions by method that were asked in Rwanda. The non-sensitive questions that were asked in the list experiments are shown in Table A4. These non-sensitive questions were selected from a larger pool of piloted questions following design advice discussed in Glynn (2013), and informed by questions used by Peterman et al. (2017) and Agüero and Frisancho (2017).

A8 show balance across the groups assigned to the list sensitive and list non-sensitive groups, and in Rwanda, also across the three groups (ACASI versus face-to-face versus list sensitive). While several of the raw differences are statistically significant at conventional levels, the differences do not appear to be meaningfully significant, they are within the bounds of what might be expected given the number of variables tested for balance, and no normalized differences are greater than 0.25 standard deviations, a suggested rule of thumb indicating good balance (Imbens and Rubin, 2015). In additional specifications, we estimate prevalence with regressions that include controls selected using post-double selection LASSO (Belloni, Chernozhukov and Hansen, 2014). This does not meaningfully change results.

4 Results

4.1 Method matters: IPV reporting differs by the survey method used

This paper assesses whether there is a difference in reported rates of intimate partner and sexual violence between the three survey methods. A difference in prevalence implies misreporting under at least one of the methods.²⁴

In Table 3, columns 1-3 show the prevalence of sexual and intimate partner violence by survey method in Rwanda and Nigeria. Columns 4-6 show the difference in prevalence between pairs of methods.²⁵ The p-values for these prevalence difference tests were adjusted for multiple hypothesis testing. Sharpened q-values were computed using the approach proposed by Benjamini and Hochberg (1995) and Anderson (2008).²⁶ Results are also shown graphically in Figure 1.

Results demonstrate that the survey method used to measure sexual and intimate partner violence can substantially affect reporting. Overall, the most widely used survey method, face-to-face interviewing, resulted in the lowest reported IPV rates, followed by ACASI, and finally the list method with the highest. Women’s reports of experiencing physical IPV in Rwanda were double when measured using the indirect, anonymous list method (20.6 percent) compared to the two direct methods (9.3 percent with face-to-face and 10.3 percent with ACASI). The differences between the list method and each of the direct methods were statistically significant, but the difference between ACASI and face-to-face prevalence was not.

In Nigeria, emotional and physical IPV rates were 33 and 39 percent greater respectively when measured using the list method (emotional violence: 39.7 percent; physical violence: 26.6 percent) compared to face-to-face (emotional violence: 29.9 percent; physical violence: 19.2 percent). The difference between list and face-to-face emotional violence prevalence was statistically significant at the 5 percent level. For physical violence, the difference was statistically significant at the 10 percent level. There was no statistically significant difference between direct and indirect rates of non-partner sexual violence in Rwanda or sexual IPV in Nigeria. Reported prevalence for these sexual violence questions was between 5.1 to 8.8 percent in Rwanda and 26 to 28.1 percent in Nigeria.

For men in Rwanda, the same pattern across methods was observed, with 13.3, 32.6, and 45.1 percent

²⁴As discussed above, we assume that in most cases, the method with the highest reported prevalence is closest to truth. This assumption is based on the results of several studies that compare observed truth with self-reported sensitive behavior, and an assumed lack of incentive to overreport IPV, particularly under the list experiment.

²⁵Non-response rates were extremely low in both contexts. For all methods, refusal rates were no greater than 0.84% for women and 2.4% for men. Non-response rates are shown in Table A16.

²⁶The q-values control for the false discovery rate for the 12 tests for difference in means that we calculate in Rwanda and for the three in Nigeria. All list method estimates were calculated with heteroskedastic-consistent robust standard errors because the error variance necessarily depends on the list group assignment (Blair and Imai, 2012).

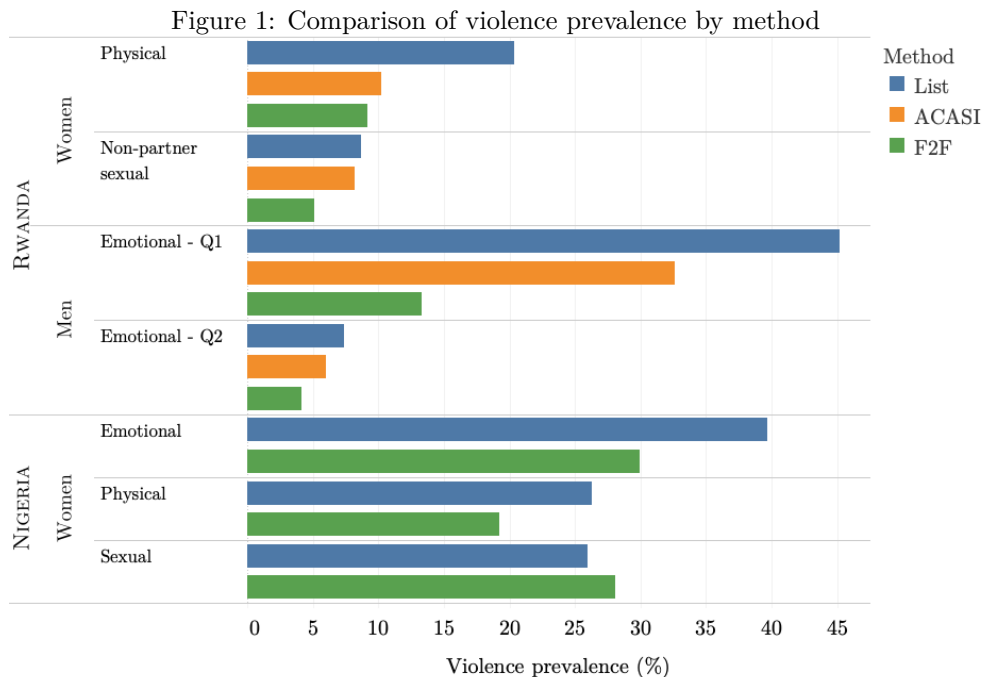
of men admitting they had tried to limit their wife’s contact with her family, with the differences all significant at the 1 percent level. These large differences between the methods suggest that, for one reason or another, men believe that they should report that they do not limit their wife’s contact with her family. The added privacy of ACASI addresses some of this underreporting, but even more is addressed by the list method’s anonymity. For the second men’s question about threatening behavior, prevalence was much lower and although the pattern held, the differences were not statistically significant.

Table 3: Prevalence of sexual and intimate partner violence by survey method

				Difference between pairs of methods		
	(1)	(2)	(3)	(4)	(5)	(6)
	Face-to-Face mean	ACASI mean	List method prevalence (difference in means)	List-F2F (3)-(1)	List-ACASI (3)-(2)	ACASI-F2F (2)-(1)
<i>Rwanda</i>						
Women						
Q1: husband pushed/ thrown you against something in past 12 months	0.093*** (0.010)	0.103*** (0.010)	0.206*** (0.035)	0.113*** [0.007]	0.103** [0.014]	0.01 [0.552]
<i>N</i>	918	914	2,728			
Q2: someone other than current husband ever forced you to have sex when didn't want to	0.051*** (0.007)	0.082*** (0.009)	0.088*** (0.034)	0.037 [0.443]	0.006 [0.872]	0.031** [0.017]
<i>N</i>	917	914	2,727			
Men						
Q1: tried to limit wife's contact with her family in past 12 months	0.133*** (0.011)	0.326*** (0.016)	0.451*** (0.038)	0.318*** [0.001]	0.125*** [0.007]	0.193*** [0.001]
<i>N</i>	909	877	2,728			
Q2: threatened to hurt wife or someone close in past 12 months	0.041*** (0.007)	0.059*** (0.008)	0.071** (0.036)	0.03 [0.542]	0.012 [0.816]	0.018 [0.113]
<i>N</i>	909	926	2,728			
<i>Nigeria</i>						
In the past 12 months, has your husband:						
Q1: said or done something to humiliate you in front of others	0.299*** (0.012)		0.397*** (0.032)	0.098** [0.015]		
<i>N</i>	1,394		2,817			
Q2: slapped you, pushed you, shaken you, or thrown something at you	0.192*** (0.011)		0.266*** (0.043)	0.074* [0.083]		
<i>N</i>	1,394		2,817			
Q3: physically forced you to have sex even when you did not want to	0.281*** (0.012)		0.260*** (0.031)	-0.017 [0.551]		
<i>N</i>	1,381		2,817			

Note: The left column shows the violence variables. Columns 1, 2 and 3 show prevalence by method. Standard errors are shown in parentheses. Columns 4, 5 and 6 shows differences in prevalence between pairs of methods. For the test of difference in means, Benjamini Hochberg (1995) adjusted q-values are displayed in square brackets, from a chi-square test of the differences in means across groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. Nigeria regressions are adjusted for sampling stratification. List estimates show robust standard errors.

Overall, these results suggest that women’s experience and men’s perpetration of emotional and physical IPV may be substantially underreported when measured using face-to-face or ACASI methods. In Rwanda, where we tested three methods, the smaller differences between ACASI and face-to-face prevalence suggest that it is primarily the list method’s anonymity rather than ACASI’s privacy that makes people more likely to report IPV. If shame were the biggest driver of misreporting, it seems likely



that we would have observed larger differences between ACASI and face-to-face prevalence, and smaller differences between list and ACASI.

If the difference in prevalence between the list and direct methods is a proxy for fear, shame or social desirability bias, then these results also suggest that questions about emotional and physical IPV are more sensitive than sexual violence, at least in Nigeria where the experiment had more statistical power. This surprising result could potentially be explained by the contexts. In Nigeria, a quarter of the women interviewed using both the list and the face-to-face methods reported experiencing sexual IPV in the past year. Qualitative interviews and focus groups suggested that there was less stigma associated with sexual IPV than physical IPV. Participants indicated that religious leaders cautioned male followers not to use physical violence against their wives, but that sexual IPV was viewed differently and was often not considered violence. Some may also consider sexual violence as resulting from a husband’s passion rather than anger, potentially making it less shameful to report than physical or emotional violence, which were sometimes seen as disciplining errant wives.

In Rwanda, however, qualitative interviews indicated that sexual violence was highly sensitive. There, the lack of difference between face-to-face and list prevalence of lifetime non-partner sexual violence may be due to truly lower prevalence and a lack of statistical power. Alternatively (or in addition), non-partner sexual violence may be so stigmatized that the mere mention of it in the list of statements is enough to make women underreport. However, given there was a difference between ACASI and face-to-face prevalence for this question, this suggests that some women are underreporting non-current partner sexual violence when interviewed face-to-face, and that the additional privacy of ACASI addresses some of their reasons for underreporting. This ACASI-face-to-face difference for the sexual violence question contrasts with the lack of difference between these methods for the physical IPV question.

Potentially, these question and method differences in Rwanda could be due to the different perceived

stakes for women who confirm having experienced IPV- which is illegal and husbands could be fined and jailed if found guilty- versus confirming experience of non-current partner sexual violence - which may feel shameful to admit face-to-face but does not carry the perceived risk that one’s husband might be jailed. This would imply that answer anonymity improves truthful reporting for sensitive questions where the perceived legal repercussions may be large, but methods like ACASI that improve privacy do not help address such misreporting. Instead, ACASI’s added privacy may primarily increase truthful reporting on sensitive questions when there is no risk of negative legal repercussions if one’s answer is discovered.

4.2 Characteristics of misreporting

In this section, we consider whether misreporting is random, and if not, identify the characteristics that predict it. To do this, we assess correlations between respondent characteristics and IPV, as reported under each of the survey methods. When the IPV-characteristic coefficient differs across methods, this implies systematic misreporting under at least one of the methods that is associated with the characteristic. This analysis provides suggestive evidence on the type of person that might misreport, potential reasons for misreporting, and the risks of drawing biased descriptive and causal inferences from such data.

In this section, we limit analysis to women’s experience of IPV given that there may be systematically different reasons for women to misreport on non-partner sexual violence questions and given that men’s self-reported perpetration is generally considered less reliable.

4.2.1 Estimation: Characteristics associated with misreporting

For those assigned to the face-to-face or ACASI groups, we assess the characteristics of those reporting IPV by regressing women’s self-reported experience of IPV against the variable of interest, such as years of education, as below.

$$Y_i = \alpha + \delta x_i + \epsilon_i$$

For the list method, we regress the number of list items reported by an individual against an indicator variable for whether they were in the list sensitive group, as well as the variable of interest and the interaction term. This regression model is shown below,

$$Y_i = \alpha + \rho T_i + \gamma x_i + \delta(T_i \cdot x_i) + \epsilon_i$$

where Y_i is the number of true statements for individual i , T_i is assignment to the sensitive list method group, and δ is the parameter of interest.²⁷ If there is no (or consistent) misreporting across the methods, the relevant coefficient for each method should be equal. If there are differences, this implies that the relationship between IPV and the characteristic is sensitive to the measurement method, and we may be estimating biased descriptive and causal relationships. For example, if we find that IPV rates are lower for educated women when we use face-to-face data, but higher for educated women when we use list method data, then we can conclude that misreporting bias is correlated with women’s education. To

²⁷This approach follows that of Holbrook and Krosnick (2010), Coutts and Jann (2011), and Agüero and Frisncho (2017). While linear regression may be less efficient than other multivariate statistical estimators proposed for list experiments by Corstange (2009), Imai (2011), Blair and Imai (2012), and Imai, Park and Greene (2015), it is simpler to interpret and compare relevant coefficients across methods and specifications for a descriptive analysis.

better understand the characteristics of those who misreport, we also compare mean IPV prevalence for respondents above and below the characteristic median (for numeric variable characteristics) or by their indicator variable characteristics.

4.2.2 Results: Respondent characteristics predict misreporting

We analyze misreporting by respondent characteristics that fall into three overlapping categories. These characteristics include: 1) those most likely to motivate respondents to misreport (to understand why people misreport IPV); 2) variables typically targeted in IPV prevention interventions (to evaluate risks of estimating biased treatment effects); and 3) characteristics frequently cited as IPV risk factors (to assess potential challenges to IPV-related theory that has been informed by analysis of face-to-face survey data). The first two categories of characteristics include perceptions of village gender norms; poverty and vulnerability; marital relationship quality; and personal gender attitudes. Additional IPV risk factors considered include women’s education and literacy, age, labor force participation and other indicators of women’s bargaining power, where theory and empirical evidence are mixed on the relationship with IPV (Bobonis, González-Brenes and Castro, 2013; Buller et al., 2018; Erten and Keskin, 2018; Heise and Kotsadam, 2015; Haushofer et al., 2019; Eswaran and Malhotra, 2011).

Tables 4 and 5 show the relationships between IPV and respondent characteristics by survey method. The first 3 (6) columns show the relevant regression coefficient of interest for Rwanda (Nigeria), and the last 3 columns show the results of chi-squared tests of the coefficient differences between pairs of methods. Each coefficient shown in the tables was estimated in a separate regression. These tables show several significant differences in coefficients across the methods, implying that a number of characteristics are associated with systematic misreporting of IPV. These results are also shown graphically in Figures A2 and A3. We also show the differences in mean IPV prevalence for respondents above and below the median of the characteristic in Tables A9 and A10 and graphically in Figures A4 and A5.

This subsection presents the results of descriptive analysis and as such, usual caution should be applied when interpreting results given that omitted variables or other sources of bias may be driving observed correlations. Given this is exploratory analysis only, we do not adjust the p-values of tests of equivalence for multiple hypothesis testing.²⁸

These results show that in both countries, vulnerable women report IPV at lower rates when they are asked about their experience using direct methods than when interviewed using the list method. Specifically, in Rwanda, a one standard deviation increase in a woman’s vulnerability index was associated with a 6 percentage point increase in reported experience of IPV when violence was measured face-to-face, a 5 percentage point increase when measured using ACASI, and a 23 percentage point increase when measured using the list method. In Nigeria, vulnerable women were also more likely to report physical IPV under the list method than face-to-face. For emotional and sexual IPV, the same pattern existed but the differences were not statistically significant. Table A9 shows that, in Rwanda, as we might expect, the difference between IPV reporting when it was measured using the list and direct methods was driven

²⁸For conceptually linked variables (for example, questions about relationship quality such as trust and empathy in the marriage), we generated weighted standardized summary indices (Anderson, 2008) to minimize the false discovery rate. For other indicators of interest where there is less theoretical coherence between the variables, we report coefficients on the raw variable. For instance, it is not clear that the following indicators of women’s bargaining power are measures of the same construct, nor whether they would each induce a respondent to underreport IPV more or less: education, asset ownership, age, marital status, and engagement in paid employment.

Table 4: Rwanda- correlations between IPV and respondent characteristics across methods

Variable	Q: husband pushed or thrown you			Method differences: P-values from tests of differences		
	(1)	(2)	(3)	(1)-(2)	(1)-(3)	(3)-(2)
	List	F2F	ACASI	List-F2F	List-ACASI	ACASI-F2F
Vulnerability index	0.23*** (0.08)	0.06*** (0.02)	0.05* (0.03)	0.07	0.05	0.65
Unequal village gender norms index	0.37*** (0.10)	0.03 (0.03)	0.04 (0.03)	0.00	0.00	0.75
Relationship quality index	0.03 (0.05)	-0.07*** (0.01)	-0.04*** (0.01)	0.06	0.17	0.18
Gender engagement index	-0.10** (0.04)	-0.02* (0.01)	-0.01 (0.01)	0.11	0.07	0.48
Progressive gender attitudes index	0.01 (0.06)	-0.01 (0.02)	-0.03 (0.02)	0.75	0.56	0.47
Bargaining power measures						
Legally married	-0.05 (0.07)	0.01 (0.02)	-0.03 (0.02)	0.49	0.83	0.21
Wife's age	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.43	0.34	0.68
Wife owns land	0.03 (0.08)	0.00 (0.02)	-0.03 (0.02)	0.72	0.49	0.36
Wife's years of education	0.02 (0.01)	-0.00 (0.00)	-0.01* (0.00)	0.13	0.06	0.27
Wife in paid work last week	-0.15** (0.07)	-0.00 (0.02)	0.03 (0.02)	0.04	0.02	0.33

Note: Standard errors are shown in parentheses (List method estimates show robust standard errors). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns 1-3 show the relevant covariate coefficients from separate regressions. Columns 4-6 show the p-values from a chi-squared test of the differences between these coefficients. The vulnerability index includes household economic shock experience, husband's trauma score, husband's alcohol consumption, and whether any household member has skipped a meal in the past week. The unequal gender norms index includes incentivized and unincentivized questions about perceived unequal gender attitudes held by others in their village. The relationship quality index includes measures of marital trust, empathy and communication. The gender engagement index includes questions on the wife's previous participation in rallies and workshops about family violence and gender equality. The gender attitudes index includes questions about respondent beliefs about gender roles and the acceptability of IPV.

by women above median vulnerability. There was no IPV reporting difference across the list and direct methods for the least vulnerable women.

More vulnerable women may face greater personal safety or economic security risks if they report that their husband perpetrates IPV and, for example, he discovered her answer or was jailed as a result. For such women, the list experiment's anonymity may provide them with additional reassurance and safety, making them more willing to report IPV. These results demonstrate that the relationship between vulnerability and IPV can be very different if estimated using anonymous list experiment IPV data. In Rwanda, it also demonstrates that ACASI's added privacy does not address the reasons vulnerable women are misreporting.

In Rwanda, we also observe a similar misreporting pattern based on a respondent's perception of unequal village gender norms, which could proxy for shame about IPV or fear of potential repercussions if a woman's survey answer were to be discovered. When a woman perceived that village gender norms were unequal, she was more likely to report IPV under the list method, where her answer is anonymous, than she was under the face-to-face or ACASI methods. Specifically, in Rwanda, a one standard deviation

Table 5: Nigeria- correlations between IPV and respondent characteristics across methods

Variable	Q1: emotional violence		Q2: physical violence		Q3: sexual violence		Method differences: P-values from tests of differences		
	(1)	(2)	(3)	(4)	(5)	(6)	(1)-(2)	(3)-(4)	(5)-(6)
	List	F2F	List	F2F	List	F2F	Q1	Q2	Q3
Vulnerability index	0.02 (0.06)	-0.08*** (0.02)	0.09 (0.07)	-0.05*** (0.02)	0.03 (0.07)	-0.05** (0.02)	0.14	0.05	0.25
Unequal village gender norms index	0.06 (0.04)	0.08*** (0.02)	0.09* (0.05)	0.05*** (0.02)	0.08* (0.04)	0.05*** (0.02)	0.61	0.43	0.50
Relationship quality index	-0.07 (0.08)	-0.21*** (0.03)	-0.02 (0.09)	-0.21*** (0.02)	-0.04 (0.08)	-0.27*** (0.03)	0.11	0.04	0.01
Public speaking confidence index	-0.02 (0.04)	-0.08*** (0.01)	0.01 (0.04)	-0.03*** (0.01)	0.00 (0.04)	-0.01 (0.01)	0.11	0.34	0.68
Progressive gender attitudes index	-0.12* (0.07)	-0.13*** (0.03)	-0.23*** (0.08)	-0.09*** (0.03)	-0.16** (0.07)	-0.16*** (0.03)	0.91	0.11	0.95
Bargaining power measures									
Wife's age	0.00 (0.00)	0.00** (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)	0.57	0.49	0.59
Wife reads and writes	0.12 (0.08)	-0.08** (0.03)	0.00 (0.10)	-0.12*** (0.02)	0.11 (0.09)	-0.04 (0.03)	0.03	0.20	0.12
Polygamous: wife number (1-4)	-0.03 (0.08)	0.02 (0.04)	0.06 (0.10)	0.03 (0.03)	0.13 (0.09)	-0.02 (0.03)	0.56	0.74	0.13
Wife brought assets to marriage	-0.04 (0.06)	-0.18*** (0.02)	0.03 (0.07)	-0.13*** (0.02)	-0.02 (0.07)	-0.09*** (0.02)	0.04	0.03	0.34

Note: Standard errors are shown in parentheses (List method estimates show robust standard errors). All regressions are adjusted for sampling stratification. *** p<0.01, ** p<0.05, * p<0.1. Columns 1-6 show the relevant covariate coefficients from separate regressions. Columns 7-9 show the p-values from a chi-squared test of the differences between these coefficients. The vulnerability index includes measures of food poverty and household economic insecurity, and wife's stress. The unequal village gender norms index includes beliefs about unequal gender attitudes held by others in the village. The relationship quality index includes measures of trust, satisfaction, empathy and communication in the marriage. The public speaking confidence index includes measures of willingness to speak up in public. The progressive gender attitudes index includes measures of personal views on the acceptability of IPV and on gender roles.

increase in a woman's perception of unequal village gender norms was associated with a 3 percentage point increase in the probability of experiencing IPV when violence was measured face-to-face, a 4 percentage point increase when measured using ACASI, and a 37 percentage point increase when using the list method. However, the positive conservative norms-IPV relationship was only statistically significant when IPV was measured using the list method. In Rwanda, the differences between the list and direct methods were statistically significant; in Nigeria, they were not.²⁹ Table A9 shows that the gender norms-IPV relationship difference between the methods was driven by the respondents who believed their village had the most conservative norms. There was no IPV reporting difference across the list and direct methods for respondents who believed their village had progressive norms. As with vulnerability, these results demonstrate that the relationship between community gender norms and IPV can be very different if estimated using anonymous list experiment data.

²⁹In Nigeria, the lack of IPV reporting difference across the methods by perceived norms may be due to the different gender norms questions that comprise the index in the two countries. The Nigerian norms index included only unincentivized questions about perceptions of other villagers' attitudes regarding the acceptability of women working or traveling outside the home. In Rwanda, the norms index focused on IPV-specific norms and included both unincentivized questions about perceptions of village IPV prevalence and others' attitudes, as well as four incentivized norms vignettes based on an adapted version of the Krupka-Weber norms coordination game (Krupka and Weber, 2013). In this game, respondents were paid if they selected the modal answer in their village about whether described behaviors about IPV and women leaving violent husbands were considered socially acceptable or unacceptable in their village. This incentive-compatible measure was intended to overcome social desirability bias.

The Rwandan result challenges an alternative hypothesis about norms and IPV misreporting which posits that in villages with highly unequal gender norms, where IPV is normalized, women may feel more able to openly report IPV. If that were the case, we should have observed no difference between the coefficients for face-to-face-estimated IPV and the two more private or anonymous methods. Instead, in Rwanda at least, it appears that only the anonymous method, but not ACASI, made vulnerable women and women living in conservative communities comfortable enough to report IPV.

Greater marital relationship quality, measured using questions about satisfaction, empathy, and trust in marriage, was also associated with underreporting. In both countries, women who reported greater relationship quality were more likely to report IPV when interviewed using the list method than face-to-face; the method differences were statistically significant on three out of the four IPV questions considered. A potential reason for women who report being in a more trusting, supportive relationship to underreport IPV when asked directly (but not indirectly) is because the direct question forces her to confront an uncomfortable truth about her marriage, or she may feel disloyal by directly admitting her husband has been violent. As Tables 4 and 5, and the associated figures A2 and A3 show, when IPV is measured using the list experiment, there is no relationship between IPV and relationship quality; it is only when using direct methods that we find a negative relationship between IPV and marital quality.

Several of these results highlight a potential measurement bias issue for the growing number of policies and interventions being assessed on their IPV impact. For example, an intervention that improves relationship quality- as several IPV prevention programs seek to do- may have no impact on actual IPV rates. However, it may lead to a reduction in a wife's willingness to directly report IPV for the reasons discussed above. This would bias treatment effect estimates, suggesting a reduction in IPV that was, in fact, an artefact of the measurement method used. Alternatively, if a policy led to a worsening in community gender norms, IPV outcomes that were measured using direct methods may not detect a resulting actual increase in IPV if women are less willing to report it. In both of these cases, list experiment-measured IPV may be less prone to reporting bias and more likely to reflect true treatment effects.

We also examined whether characteristics that are proposed as causes or common correlates of IPV remain stable or at least with the same sign on the relationship, regardless of the method used, or whether the nature of the relationship depends on the measurement method. Across different indicators of women's intrahousehold bargaining power, we find some evidence of systematic misreporting of IPV, though no clear pattern across different indicators. For example, in Rwanda, we find a strong relationship between employment and IPV misreporting. Results in Table A9 show that this reporting effect is driven by a large difference in reporting for women who are not engaged in paid employment. Similar to the norms and vulnerability results, this may suggest that women with the weakest bargaining power or outside options underreport the most when asked directly about their IPV experience. Again, the list experiment appears to make these women more willing to report IPV.

Potentially most concerning, we observe some opposite-signed relationships, where the sign on the relationship between IPV and the respondent's characteristic is positive for one measurement method and negative for another method. In Rwanda, these variables include wife's years of education, relationship quality, and engagement in paid work, and in Nigeria, wife's vulnerability, literacy, and assets brought into the marriage. For example, in both Rwanda and Nigeria, there is suggestive evidence of a positive relationship between IPV and women's education when IPV is measured directly, but a negative

relationship when using the list method. Specifically, literate Nigerian women are 12 percentage points more likely to experience emotional IPV if it is measured using the list method, and 8 percentage points less likely if IPV is measured face-to-face. This is similar to results in Peru and India, where researchers found women’s education to be positively correlated with IPV when measured using the list method but negatively when measured face-to-face (Agüero and Frisanco, 2017; Joseph et al., 2017). The opposite-signed IPV-education associations imply opposing theories about women’s bargaining power as a cause of IPV, highlighting the importance of minimizing measurement bias to better understand IPV causes, theory, and prevention.

Overall, the sensitivity of the relationships between IPV and key bargaining power and other indicators to the survey method used is concerning. It suggests that caution should be applied when using direct IPV measures to understand descriptive and causal relationships as misreporting may bias correlations and treatment effect estimates. The anonymity of the list experiment may minimize this reporting bias.

4.3 Robustness checks

In the section below, we report the results of robustness checks on the list method, followed by checks on the overall analysis.

4.3.1 Checks on list method assumptions

There are two assumptions that underpin the validity of statistical analyses of list experiments, and which, if they hold, should result in an unbiased difference-in-means estimator (Blair and Imai, 2012). The first assumption is that of ‘no design effect’, meaning that the inclusion of the sensitive item in the list does not affect answers to non-sensitive items. The second assumption is that of ‘no liars’, or that the answers given for the sensitive item are truthful.

Blair and Imai (2012) propose a likelihood ratio test to assess the probability that the first assumption is violated. The test is based on the assumption that the inclusion of a sensitive statement should not change the sum of affirmative answers to the non-sensitive statements, and that cumulative proportions of different respondent types should be non-negative (Bulte and Lensink, 2019). If one of the proportions is negative, then the no design effect assumption has likely been violated. The rejection of the null hypothesis of no design effect using Blair and Imai’s statistical test, made available on the open-source software R (Blair and Imai, 2011), would suggest there was a design effect. The test returns a Bonferroni-corrected minimum p-value of the test.

In Nigeria, the test returned a minimum p-value of 1, suggesting there was no design effect for the three list questions. In Rwanda, where the list non-sensitive group was directly asked each non-sensitive question, there is statistically significant evidence of a design effect for three out of the four questions, suggesting that group assignment affected answers to the non-sensitive questions. There was no design effect for the first men’s question.

What then is the potential direction of bias introduced by this assumption violation in Rwanda? In the three cases in Rwanda where this assumption was violated, the design effect is likely to have resulted in downwards-biased prevalence estimates for the list method. This is because the tests that it failed, assessing whether the cumulative proportion of responses in the non-sensitive group was greater than the proportion in the sensitive list group, were for the share of respondents reporting that zero items were true

for them. More respondents in the list sensitive group reported zero than in the non-sensitive group. This implies that the mean in the list sensitive group is likely to be downwards biased, implying the same for the difference-in-means estimate and the resulting IPV prevalence.³⁰ Further, the one Rwanda question that did not fail the no design effects test showed the greatest difference in prevalence between the list and other methods. It therefore seems plausible that, if biased, the three Rwandan list experiments that failed the no design effects test produced downwards biased prevalence estimates. The associated test p-values and proportions are shown in Tables A11 and A12.

There is no explicit test for the second ‘no liars’ assumption underpinning the list experiment. However, the most obvious threat to this assumption is when a respondent’s anonymity for their answer to the sensitive statement is compromised. This ‘ceiling effect’ exists when a respondent’s true answer is affirmative for all of the statements, thus their affirmative response to the sensitive statement will no longer be hidden. However, the proportion of respondents answering with the maximum number of items (three and four for the non-sensitive and sensitive list groups, respectively) is within the range of other studies.³¹ Further, the presence of a ceiling effect likely indicates downwards biased estimates because other respondents would be likely to have understated their true number of 4 affirmative answers in order to conceal their affirmative sensitive answer (Blair and Imai, 2012).

4.3.2 Checks on overall approach

The two key assumptions underpinning the unbiasedness of the list experiment estimator have parallels for the experiment as a whole. While there is no direct test for these assumptions holding for the ACASI and face-to-face methods, there are some indicators of whether or not they were violated.

First, we checked balance on survey questions asked of all respondents after the experimental violence questions. If respondents were differentially affected by one survey method, perhaps indicating a design effect, we may observe them answering subsequent survey questions differently. We found very few statistically significant differences in respondent answers despite the many tests for significance. None of these differences are materially significant.

Second, we analyzed the main list method prevalence regressions with enumerator fixed effects and controls selected using post-double-selection LASSO (Belloni, Chernozhukov and Hansen, 2014). We found no notable differences in coefficients or significance. These results are shown in Tables A13 to A15.

Third, although a general ‘no liars’ assumption is untestable, we compared refusal rates across methods. This could indicate the relative risk of misreporting for each method. The refusal rate with ACASI, where all one has to do to skip a question is touch a button, may be a lower-bound on the share of people who might have lied if they were interviewed face-to-face and would be uncomfortable asking an

³⁰One study by Flavin and Keane (2009) evaluates whether the design of the non-sensitive list affects responses. This study found that asking the non-sensitive group directly about the non-sensitive items resulted in a higher mean number of items than when the non-sensitive questions were asked using the traditional ‘how many of the following items’ approach. As a result, the direct approach to asking the non-sensitive questions resulted in downwards-biased prevalence estimates, though it should be noted that their ‘non-sensitive’ items were somewhat sensitive. Further, a recent paper by Chuang et al. (2018) found that having relatively sensitive ‘non-sensitive’ statements on the list appears to mask the sensitive item of interest, and results in lower variance and higher prevalence estimates. In Rwanda and Nigeria, we used relatively innocuous non-sensitive list items, which, if results from the Chuang et al. measurement experiment on risky sexual behavior hold, then it again seems plausible that these list experiment results may be downwards biased.

³¹The proportion of respondents in the non-sensitive list group reporting three items as true is never greater than 15.3 percent (for the women’s physical violence question in Rwanda), and the proportion of respondents in the sensitive list group reporting the full number of items is never greater than 5.4 percent (for the women’s physical violence question in Nigeria).

enumerator to skip the question. Indeed, we find that ACASI, has the highest refusal rate out of the methods, though it is still low, at 0.84 percent for women. Table A16 shows refusal rates by question and method.

Additional indications that these results are robust are supported by the broadly similar patterns of misreporting by survey method that were observed over different countries, genders, violence types, perpetration and victimization questions, and list method designs.

5 Conclusion

This paper provided some of the first evidence that the prevalence of intimate partner violence varies substantially depending on the survey method used to measure it. Our results suggest that in some contexts, standard survey methods are likely to result in significant underestimates of the prevalence and welfare costs of IPV. The most widely used method, face-to-face interview, resulted in the lowest prevalence estimates, followed by the more private ACASI method. Finally, the list method, which allows respondents to report anonymously, resulted in the highest prevalence. In Rwanda, physical IPV rates were double, and in Nigeria, 39 percent greater when measured using the list method compared to the direct methods.

This paper also found that misreporting was systematic. The women who underreported when interviewed directly (compared to when interviewed with the list method) tended to be more vulnerable, not engaged in paid work, living in conservative communities, and more educated. These results demonstrate that misreporting was correlated with indicators that theory suggests are risk factors for IPV and with variables often targeted in women’s empowerment programs. Our results show non-trivial patterns of misreporting, such that gradients on some IPV-predictor relationships changed substantially, and some signs flipped, depending on the method used to measure IPV.

As other studies have found, direct measures of IPV can produce biased treatment effect estimates, for example, finding programs caused a reduction in IPV if measured face-to-face, but an increase if measured with the list experiment (Bulte and Lensink, 2019). Given the growing number of studies measuring program impacts on IPV, our results further highlight the risks of collecting measurement error-prone data for generating unbiased correlations and treatment effect estimates. Given that most IPV-related theory and prevention interventions are informed by face-to-face survey data, misreporting bias could have significant implications for our understanding of the prevalence of IPV, what causes it, and how to prevent it.

Our three-method comparison in Rwanda provides additional insights into potential reasons for misreporting. There were strikingly few differences between ACASI-measured and face-to-face-measured prevalence and IPV-predictor relationships. This suggests that women in Rwanda were no more comfortable reporting IPV with ACASI’s added privacy, yet the list method’s anonymity made many more women willing to report. This may suggest that on questions about illegal behaviors where respondents fear a risk of legal sanctions or repercussions from their answer, for example, in places where physical and sexual IPV are criminalized, then only anonymous survey methods may induce some respondents to report.³²

³²The similar face-to-face and ACASI reporting patterns may also be affected by the strict private interview setting in Rwanda. In a busier, more typical survey setting where privacy from curious onlookers and other household members is harder to enforce, we may have found that women would be even less willing to report IPV face-to-face, and ACASI may

Our results also have implications for the growing literature that uses face-to-face IPV data from multiple countries to draw inferences about IPV prevalence, predictors, and causes (Devries et al., 2013; García-Moreno et al., 2013; Cools and Kotsadam, 2017; Jewkes et al., 2017; Alesina, Brioschi and La Ferrara, 2020; Heise and Kotsadam, 2015). Our finding of significant and systematic underreporting for some types of IPV in our two rural African contexts contrasts with findings from Lima, Peru, where researchers found no overall difference between face-to-face and list method prevalence (Agüero and Frisancho, 2017). These divergent results indicate that reporting bias is, unsurprisingly, strongly affected by contextual factors. Given significant, systematic face-to-face misreporting in some contexts but not others, our results imply that caution should be used when aggregating and analyzing face-to-face IPV data from diverse settings, lest context-specific misreporting bias results.

Policy makers and social scientists measuring sensitive topics could use similar randomized survey experiments to compare IPV reporting rates between the list experiment and direct methods to identify levels of stigma, the types of people most likely to misreport, and the risk of generating biased treatment effects. For example, social scientists researching the impact of interventions on IPV might use such an experiment to sign potential reporting bias in their treatment effect estimates. Also, given Demographic and Health Surveys' large sample sizes and use in multi-country analysis, perhaps such surveys should measure several IPV questions using the list method as well as the face-to-face method. This would allow data users to assess the likely extent of reporting bias by context, allowing them to make adjustments to inferences about true levels and correlates of IPV.³³

Future research could be conducted in other contexts and on other types of violence and sensitive topics to better understand what drives misreporting. Other potential areas for additional research include studying how to best design and administer the list experiment to minimize bias, and developing alternative measurement methods for efficient causal research on IPV which, according to face-to-face survey data, is estimated to affect one in three women globally (García-Moreno et al., 2013).

have made more of a difference. See Appendix B for a further discussion on IPV measurement method trade-offs.

³³DHS data on IPV cover more than 60 developing countries and are cited in more than 5,000 papers on google scholar. If DHS IPV data are measured with substantial, systematic, and context-specific bias as this paper suggests is likely, this has significant implications for the use of such data.

References

- Agüero, Jorge, and Verónica Frisancho.** 2017. “Misreporting in Sensitive Health Behaviors and Its Impact on Treatment Effects: An Application to Intimate Partner Violence.” IDB Working Paper No. IDB-WP-853.
- Agüero, Jorge, Úrsula Aldana, Erica Field, Verónica Frisancho, and Javier Romero.** 2020. “Is Community-Based Targeting Effective in Identifying Intimate Partner Violence?” *AEA Papers and Proceedings*, 110(May): 605–09.
- Alesina, Alberto, Benedetta Brioschi, and Eliana La Ferrara.** 2020. “Violence Against Women: A Cross-cultural Analysis for Africa.” *Economica*. May.
- Alik-Lagrange, Arthur, Claire Cullen, Mũthoni Ngatia, and Julia Vaillant.** 2017. “Preventing intimate partner violence: Impact Evaluation of MIGEPROF couples training for IPV prevention in Eastern Rwanda.” AEA RCT Registry. June 28. <https://doi.org/10.1257/rct.2282-2.0>.
- Anderson, Michael L.** 2008. “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of the American Statistical Association*, 103(484): 1481–1495.
- Banerjee, Abhijit, Eliana La Ferrara, and Victor Orozco.** 2019. “Entertainment, education, and attitudes toward domestic violence.” *AEA Papers and Proceedings*, 109: 133–37.
- Bastian, Gautam Gustav, Markus P. Goldstein, and Sreelakshmi Papineni.** 2017. “Are cash transfers better chunky or smooth?: evidence from an impact evaluation of a cash transfer program in northern Nigeria.” World Bank Gender Innovation Lab Policy Brief, Washington, D.C.
- Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen.** 2014. “Inference on treatment effects after selection among high-dimensional controls.” *The Review of Economic Studies*, 81(2): 608–650.
- Benjamini, Yoav, and Yosef Hochberg.** 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1): 289–300.
- Bharadwaj, Prashant, Mallesh M. Pai, and Agne Suziedelyte.** 2017. “Mental health stigma.” *Economics Letters*, 159: 57–60.
- Blair, Graeme, Alex Coppock, and Margaret Moor.** 2018. “When to Worry About Sensitivity Bias: Evidence from 30 Years of List Experiments.” Retrieved from: <https://graemeblair.com/papers/sensitivity-bias.pdf>. Accessed on 15 May 2020.
- Blair, Graeme, and Kosuke Imai.** 2011. “List: Statistical Methods for the Item Count Technique and List Experiment. Comprehensive R Archive Network (CRAN).”
- Blair, Graeme, and Kosuke Imai.** 2012. “Statistical Analysis of List Experiments.” *Political Analysis*, 20: 47–77.

- Blair, Graeme, Kosuke Imai, and Jason Lyall.** 2014. “Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan.” *American Journal of Political Science*, 58(4): 1043–1063.
- Blattman, Christopher, Julian Jamison, Tricia Koroknay-Palicz, Katherine Rodrigues, and Margaret Sheridan.** 2016. “Measuring the measurement error: A method to qualitatively validate survey data.” *Journal of Development Economics*, 120: 99–112.
- Bobonis, Gustavo J, Melissa González-Brenes, and Roberto Castro.** 2013. “Public Transfers and Domestic Violence: The Roles of Private Information and Spousal Control.” *American Economic Journal: Economic Policy*, 5(51): 179–205.
- Buller, Ana Maria, Amber Peterman, Meghna Ranganathan, Alexandra Bleile, Melissa Hidrobo, and Lori Heise.** 2018. “A Mixed-Method Review of Cash Transfers and Intimate Partner Violence in Low- and Middle-Income Countries.” *The World Bank Research Observer*, 33(2): 218–258.
- Bulte, Erwin, and Robert Lensink.** 2019. “Women’s empowerment and domestic abuse: Experimental evidence from Vietnam.” *European Economic Review*, 115: 172–191.
- Chuang, Erica, Pascaline Dupas, Elise Huillery, and Juliette Seban.** 2018. “Sex, Lies, and Measurement.” Retrieved from: https://web.stanford.edu/~pdupas/CDHS_measurement.pdf. Accessed on 1 May 2019.
- Cools, Sara, and Andreas Kotsadam.** 2017. “Resources and intimate partner violence in Sub-Saharan Africa.” *World Development*, 95: 211–230.
- Corstange, Daniel.** 2009. “Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT.” *Political Analysis*, 17(1): 45–63.
- Coutts, Elisabeth, and Ben Jann.** 2011. “Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT).” *Sociological Methods & Research*, 40(1): 169–193.
- De Cao, Elisabetta, and Clemens Lutz.** 2018. “Sensitive Survey Questions: Measuring Attitudes Regarding Female Genital Cutting Through a List Experiment.” *Oxford Bulletin of Economics and Statistics*, 80(5): 871–892.
- Devries, Karen M, Joelle YT Mak, Claudia García-Moreno, Max Petzold, James C Child, Gail Falder, Stephen Lim, Loraine J Bacchus, Rebecca E Engell, Lisa Rosenfeld, et al.** 2013. “The global prevalence of intimate partner violence against women.” *Science*, 340(6140): 1527–1528.
- DHS.** 2014. “Nigeria Demographic and Health Survey.” National Population Commission - NPC/Nigeria and ICF International, Abuja.
- DHS.** 2016. “Rwanda Demographic and Health Survey 2014-15.” National Institute of Statistics of Rwanda (NISR), Ministry of Health (MOH), and ICF International, Kigali and Maryland.

- Erten, Bilge, and Pinar Keskin.** 2018. “For Better or for Worse?: Education and the Prevalence of Domestic Violence in Turkey.” *American Economic Journal: Applied Economics*, 10(1): 64–105.
- Eswaran, Mukesh, and Nisha Malhotra.** 2011. “Domestic violence and women’s autonomy in developing countries: theory and evidence.” *Canadian Journal of Economics/Revue canadienne d’économique*, 44(4): 1222–1263.
- Falb, Kathryn, Sophie Tanner, Khudejha Asghar, Samir Souidi, Stan Mierzwa, Asham Assazenew, Theresita Bakomere, Pamela Malinga, Katie Robinette, Woinishet Tibebu, and Lindsay Stark.** 2016. “Implementation of Audio-Computer Assisted Self-Interview (ACASI) among adolescent girls in humanitarian settings: feasibility, acceptability, and lessons learned.” *Conflict and Health*, 10(1): 32.
- Fay, R E, C F Turner, A D Klassen, J H Gagnon, J. H. Pleck, and F. L. Sonenstein.** 1998. “Prevalence and patterns of same-gender sexual contact among men.” *Science*, 280(5365): 867–873.
- Fincher, Danielle, Kristin VanderEnde, Kia Colbert, Debra Houry, L. Shakiyla Smith, and Kathryn 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): 818–838.
- Flavin, Patrick, and Michael Keane.** 2009. “How Angry am I? Let Me Count the Ways: Question Format Bias in List Experiments.”
- García-Moreno, C, C. Pallitto, K. Devries, H. Stöckl, 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.” Geneva:World Health Organization.
- Global Burden of Disease.** 2017. “Global Burden of Disease Study 2016: Health-related Sustainable Development Goals (SDG) Indicators 1990-2030.” Seattle, United States: Institute for Health Metrics and Evaluation (IHME).
- Glynn, Adam N.** 2013. “What Can We Learn with Statistical Truth Serum?: Design and Analysis of the List Experiment.” *Public Opinion Quarterly*, 77(S1): 159–172.
- Gonzalez-Ocantos, Ezequiel, Chad Kiewiet de Jonge, Carlos Meléndez, Javier Osorio, and David W. Nickerson.** 2012. “Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua.” *American Journal of Political Science*, 56(1): 202–217.
- Gottschalk, Peter, and Minh Huynh.** 2010. “Are Earnings Inequality and Mobility Overstated? The Impact of Nonclassical Measurement Error.” *Review of Economics and Statistics*, 92(2): 302–315.
- Haushofer, Johannes, Charlotte Ringdal, Jeremy P Shapiro, and Xiao Yu Wang.** 2019. “Income changes and intimate partner violence: Evidence from unconditional cash transfers in Kenya.” NBER Working Paper No. 25627.
- Heise, Lori, and Andreas Kotsadam.** 2015. “Cross-national and multilevel correlates of partner violence: an analysis of data from population-based surveys.” *The Lancet Global Health*, 3(6): 332–340.

- Heise, Lori, and Mazedra Hossain.** 2017. “Measuring intimate partner violence.” STRIVE, London.
- Holbrook, A. L., and J. A. Krosnick.** 2010. “Measuring Voter Turnout By Using The Randomized Response Technique: Evidence Calling Into Question The Method’s Validity.” *Public Opinion Quarterly*, 74(2): 328–343.
- Imai, Kosuke.** 2011. “Multivariate Regression Analysis for the Item Count Technique.” *Journal of the American Statistical Association*, 106(494): 407–416.
- Imai, Kosuke, Bethany Park, and Kenneth F Greene.** 2015. “Using the Predicted Responses from List Experiments as Explanatory Variables in Regression Models.” *Political Analysis*, 23: 180–196.
- Imbens, Guido., and Donald B. Rubin.** 2015. *Causal inference for statistics, social, and biomedical sciences : an introduction*. Cambridge University Press.
- Jamison, Julian C, Dean Karlan, and Pia Raffler.** 2013. “Mixed method evaluation of a passive mHealth sexual information texting service in Uganda.” NBER Working Paper No. 19107.
- Jarlais, Don C Des, Denise Paone, Judith Milliken, Charles F Turner, Heather Miller, James Gribble, Qiuhu Shi, Holly Hagan, and Samuel R Friedman.** 1999. “Audio-computer interviewing to measure risk behaviour for HIV among injecting drug users: a quasi-randomised trial.” *The Lancet*, 353(9165): 1657–1661.
- Jaya, Michelle J Hindin, and Saifuddin Ahmed.** 2008. “Differences in young people’s reports of sexual behaviors according to interview methodology: a randomized trial in India.” *American Journal of Public Health*, 98(1): 169–74.
- Jewkes, Rachel, Emma Fulu, Ruchira Tabassam Naved, Esnat Chirwa, Kristin Dunkle, Regine Haardörfer, Claudia Garcia-Moreno, et al.** 2017. “Women’s and men’s reports of past-year prevalence of intimate partner violence and rape and women’s risk factors for intimate partner violence: A multicountry cross-sectional study in Asia and the Pacific.” *PLoS medicine*, 14(9).
- Joseph, George, Syed Usman Javaid, Luis Alberto Andres, Gnanaraj Chellaraj, Jennifer L Solotaroff, and S Irudaya Rajan.** 2017. “Underreporting of gender-based violence in Kerala, India: an application of the list randomization method.” World Bank Policy Research Working Paper 8044.
- Karlan, Dean S, and Jonathan Zinman.** 2012. “List randomization for sensitive behavior: An application for measuring use of loan proceeds.” *Journal of Development Economics*, 98(1): 71–75.
- Krupka, EL, and RA Weber.** 2013. “Identifying social norms using coordination games: Why does dictator game sharing vary?” *Journal of the European Economic Association*, 11(3): 495–524.
- Kuklinski, James H., Michael D. Cobb, and Martin Gilens.** 1997. “Racial Attitudes and the New South.” *The Journal of Politics*, 59(2): 323–349.
- Langhaug, Lisa F., Lorraine Sherr, and Frances M. Cowan.** 2010. “How to improve the validity of sexual behaviour reporting: systematic review of questionnaire delivery modes in developing countries.” *Tropical Medicine & International Health*, 15(3): 362–381.

- Lépine, Aurélia, Carole Treibich, Cheikh Tidiane Ndour, Khady Gueye, and Peter Vickerman.** 2020. “HIV infection risk and condom use among sex workers in Senegal: evidence from the list experiment method.” *Health Policy and Planning*, 35(4): 408–415.
- Mensch, Barbara S., Paul C. Hewett, Richard Gregory, and Stephane Helleringer.** 2008. “Sexual Behavior and STI/HIV Status Among Adolescents in Rural Malawi: An Evaluation of the Effect of Interview Mode on Reporting.” *Studies in Family Planning*, 39(4): 321–334.
- Meyer, Bruce D., and Nikolas Mittag.** 2019. “Using linked survey and administrative data to better measure income: implications for poverty, program effectiveness, and holes in the safety net.” *American Economic Journal: Applied Economics*, 11(2): 176–204.
- Miller, Judith D.** 1984. “A New Survey Technique For Studying Deviant Behavior.” PhD diss. George Washington University.
- Murray-Close, M., and M.L. Heggeness.** 2018. “Manning up and womaning down: How husbands and wives report their earnings when she earns more.” US Census Bureau Social, Economic, and Housing Statistics Division Working Paper.
- Palermo, T., J. Bleck, and A. Peterman.** 2014. “Tip of the Iceberg: Reporting and Gender-Based Violence in Developing Countries.” *American Journal of Epidemiology*, 179(5): 602–612.
- Peterman, Amber, Tia M. Palermo, Sudhanshu Handa, and David Seidenfeld.** 2017. “List randomization for soliciting experience of intimate partner violence: Application to the evaluation of Zambia’s unconditional child grant program.” *Health Economics*, 27(3): 622–628.
- Phillips, Anna E, Gabriella B Gomez, Marie-Claude Boily, and Geoffrey P Garnett.** 2010. “A systematic review and meta-analysis of quantitative interviewing tools to investigate self-reported HIV and STI associated behaviours in low- and middle-income countries.” *International Journal of Epidemiology*, 39(6): 1541–1555.
- Rosenfeld, Bryn, Kosuke Imai, and Jacob N. Shapiro.** 2016. “An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions.” *American Journal of Political Science*, 60(3): 783–802.
- Tourangeau, Roger, and Ting Yan.** 2007. “Sensitive questions in surveys.” *Psychological Bulletin*, 133(5): 859–883.
- Traunmüller, Richard, Sara Kijewski, and Markus Freitag.** 2019. “The Silent Victims of Sexual Violence during War: Evidence from a List Experiment in Sri Lanka.” *Journal of Conflict Resolution*, 63(9): 2015–2042.
- WHO.** 2001. “Putting women’s safety first: Ethical and safety recommendations for research on domestic violence against women.” World Health Organization, Geneva.
- WHO.** 2016. “Ethical and safety recommendations for intervention research on violence against women.” World Health Organization, Geneva.

World Bank. 2017. "Experimental Evaluation of MTV Shuga: Changing social norms and behaviours with entertainment education." World Bank Development Impact Evaluation (DIME). Retrieved from: <http://pubdocs.worldbank.org/en/438421467236362785/Entertainment-Edu-workshop-Flyer-6-3-16.pdf>.

A Appendix: Figures and Tables

Figure A1: ACASI screen and administration



Table A1: Rwanda: survey structure

Consent form		
Demographics		
Housing, assets & income		
Labour & agriculture		
Household decision-making		
Gender attitudes & relationship quality		
Personal attitudes about IPV (ACASI)		
Experience of IPV (ACASI)- women; Likelihood of perpetrating violence (ACASI)- men		
2 experimental sexual violence + IPV questions		
<i>ACASI</i>	<i>Face-to-face</i>	<i>List method</i>
2 direct ACASI questions	2 direct face-to-face questions	2 sensitive list questions
6 direct non-sensitive questions	6 direct non-sensitive questions	
Beliefs about community IPV attitudes and behaviour		
Social networks		
Behavioural games		

Table A2: Nigeria: survey structure

Consent form
Household roster- demographics
Business & income
Labour
Expenditure
Assets & housing
Nutrition & food security
Plot & crop roster
Happiness, shocks & family history
Savings and loans
Gender attitudes & relationship quality
3 experimental IPV questions
<i>Face-to-face</i>
<i>List method</i>
3 non-sensitive sets of list questions
3 sensitive sets of list questions
Women's empowerment in agriculture
3 direct experimental questions
5 direct IPV questions
Anthropometrics measures

Table A3: Rwanda- physical violence question wording by method

Sensitive list group	Non-sensitive list method group	
List experiment	Face-to-face	ACASI
Of the following 4 things, how many HOW MANY of them have happened to you in the past 12 months?	1. In the past 12 months, has your partner pushed or thrown you against something like a wall or piece of furniture?	1. If your partner has pushed or thrown you against something like a wall or piece of furniture in the past 12 months, press the blue square. If he has not, press the red square. If you do not want to answer this question, press the star.
1. You have taken care of a sick family member	2. Have you taken care of a sick family member in the past 12 months?	2. Have you taken care of a sick family member in the past 12 months?
2. You have gone to a health worker to talk about your health	3. Have you gone to a health worker to talk about your health in the past 12 months?	3. Have you gone to a health worker to talk about your health in the past 12 months?
3. You forgot to return a borrowed item	4. Have you forgotten to return a borrowed item in the past 12 months?	4. Have you forgotten to return a borrowed item in the past 12 months?
4. Your husband has pushed or thrown you against something like a wall or piece of furniture.		
[0,1,2,3, or 4]	[yes/no]	[yes/no]

Table A4: Non-sensitive questions

Rwanda - Has this happened to you in the past 12 months/ ever?

Women Q1

Have you taken care of a sick family member in the past 12 months?
 Have you gone to a health worker and talked about your health in the past 12 months?
 Have you forgotten to return a borrowed item in the past 12 months?

Women Q2

Have you ever borrowed a neighbor's tool for farm work?
 Have you ever travelled to another village for a week or more for work?
 Have you ever used a phone to transfer mobile phone credit to a friend?

Men Q1

Have you gone to a health worker and talked about your health in the past 12 months?
 Have you forgotten to return a borrowed item in the past 12 months?
 Have you talked to a friend about their health troubles in the past 12 months?

Men Q2

Have you borrowed a neighbor's tool for farm work in the past 12 months?
 Have you traveled to another village for work for a week or more in the past 12 months?
 Have you used a phone to transfer mobile phone credit to a friend in the past 12 months?

Nigeria - HOW MANY of the following 3 things have happened to you in the past 12 months?

Q1

You purchased a fan
 You went to Abuja with a family relative
 You participated in a marriage celebration in a neighborhood household

Q2

You gave a gift to another woman who gave birth in the village
 You talked to a health worker about your health
 You talked to the village leader

Q3

You participated in a name giving ceremony in a neighborhood household
 You went to Kano
 You took care of a sick family member who was unable to care for themselves

Figure A2: Rwanda- relationships between physical IPV and respondent characteristics across survey methods

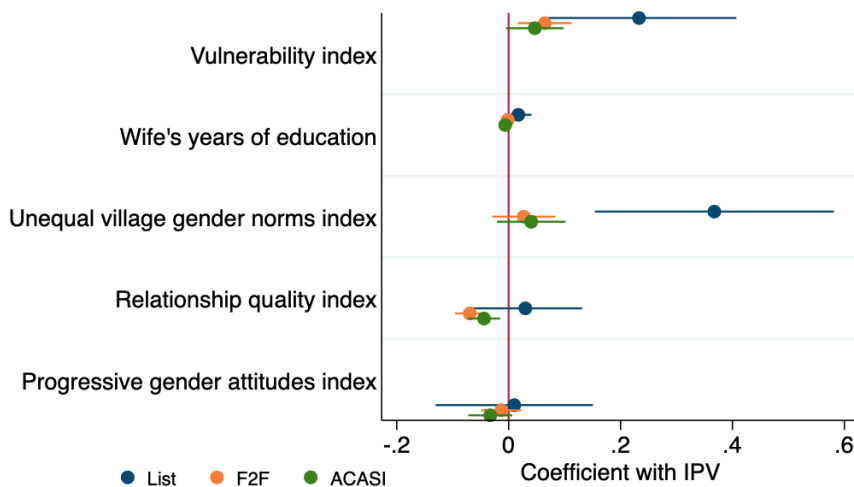


Table A5: Rwanda: women- balance across 3 survey method groups

Variable	(1) ACASI		(2) Face-to-face		(3) List		(4) (5) (6) Normalized difference			F-test for joint orthogonality
	N	Mean/SE	N	Mean/SE	N	Mean/SE	(1)-(2)	(1)-(3)	(2)-(3)	
Age	935	36.858 [0.309]	919	37.513 [0.329]	870	36.733 [0.335]	-0.07	0.01	0.08	0.189
Age at marriage/living with partner	934	21.301 [0.134]	918	21.788 [0.158]	870	21.320 [0.145]	-0.109	0.00	0.10	0.03**
In a civil/legal marriage	935	0.491 [0.016]	919	0.491 [0.016]	870	0.500 [0.017]	0.00	-0.02	-0.02	0.91
Wife's years of education	935	4.443 [0.102]	919	4.405 [0.098]	870	4.199 [0.097]	0.01	0.08	0.07	0.18
Number of biological children	935	3.775 [0.072]	919	3.804 [0.073]	869	3.776 [0.076]	-0.01	0.00	0.01	0.951
Wife worked as a paid employee in past week	935	0.391 [0.016]	919	0.402 [0.016]	870	0.399 [0.017]	-0.02	-0.02	0.01	0.9
Savings (RWF)	935	23433.950 [1363.894]	919	29867.790 [5617.442]	870	25093.729 [2741.997]	-0.05	-0.026	0.04	0.438
One spouse is a member of a VSLA	935	0.748 [0.014]	919	0.749 [0.014]	870	0.748 [0.015]	0.00	0.00	0.00	0.999
Wife worked in agriculture in past week	935	0.898 [0.010]	919	0.899 [0.010]	870	0.891 [0.011]	0.00	0.03	0.03	0.824
Husband ever drinks alcohol	932	0.572 [0.016]	915	0.532 [0.017]	867	0.602 [0.017]	0.08	-0.06	-0.14	0.01**
Husband has other wives/partners	934	0.040 [0.006]	918	0.035 [0.006]	867	0.040 [0.007]	0.03	0.00	-0.03	0.803

Notes: The value displayed for F tests are F statistics for the differences in means across groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A6: Rwanda: women- balance across 2 list method groups

Variable	(1) List non-sensitive group		(2) List sensitive group		(3) Normalized difference	(4) F-test for joint orthogonality
	N	Mean/SE	N	Mean/SE	(1)-(2)	
Age	1854	37.182 [0.226]	870	36.733 [0.335]	0.046	0.264
Age at marriage/living with partner	1852	21.542 [0.104]	870	21.320 [0.145]	0.051	0.219
In a civil/legal marriage	1854	0.491 [0.012]	870	0.500 [0.017]	-0.018	0.656
Wife's years of education	1854	4.424 [0.071]	870	4.199 [0.097]	0.075	0.067*
Number of biological children	1854	3.790 [0.052]	869	3.776 [0.076]	0.006	0.878
Wife worked as a paid employee in past week	1854	0.396 [0.011]	870	0.399 [0.017]	-0.005	0.905
Savings (RWF)	1854	26623.108 [2868.367]	870	25093.729 [2741.997]	0.014	0.739
One spouse is a member of a VSLA	1854	0.748 [0.010]	870	0.748 [0.015]	-0.000	0.993
Wife worked in agriculture in past week	1854	0.899 [0.007]	870	0.891 [0.011]	0.026	0.534
Husband ever drinks alcohol	1847	0.552 [0.012]	867	0.602 [0.017]	-0.101	0.015**
Husband has other wives/partners	1852	0.037 [0.004]	867	0.040 [0.007]	-0.016	0.693

Notes: The value displayed for F tests are F statistics for the differences in means across groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A7: Rwanda: men- balance across 3 survey method groups

Variable	(1)		(2)		(3)		Normalized difference			F-test for joint orthogonality
	N	Mean/SE	N	Face-to-face Mean/SE	N	List Mean/SE	(1)-(2)	(1)-(3)	(2)-(3)	
Age	940	40.895 [0.351]	907	41.183 [0.348]	877	40.958 [0.357]	-0.03	-0.006	0.02	0.83
Age at marriage/living with partner	940	25.367 [0.188]	907	25.696 [0.207]	877	25.613 [0.206]	-0.06	-0.04	0.01	0.48
Household size (excluding self and partner)	940	3.491 [0.063]	907	3.492 [0.066]	877	3.486 [0.063]	-0.000	0.00	0.00	1.00
Husband's years of education	940	4.566 [0.096]	907	4.605 [0.106]	877	4.420 [0.106]	-0.01	0.05	0.06	0.41
Household member skipped meal last week	940	0.417 [0.016]	907	0.396 [0.016]	877	0.431 [0.017]	0.04	-0.03	-0.07	0.31
Paid bride price	940	0.373 [0.016]	907	0.399 [0.016]	877	0.391 [0.016]	-0.05	-0.04	0.02	0.51
Worked as paid employee in past week	940	0.583 [0.016]	907	0.579 [0.016]	877	0.577 [0.017]	0.01	0.01	0.00	0.97
Number of traumatic experiences in conflict (max 7)	939	3.265 [0.050]	905	3.255 [0.050]	877	3.282 [0.052]	0.01	-0.01	-0.02	0.93
Contraception use	938	0.755 [0.014]	907	0.766 [0.014]	875	0.759 [0.014]	-0.03	-0.01	0.02	0.84

Notes: The value displayed for F tests are F statistics for the differences in means across groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A8: Nigeria: balance across 2 survey method groups

Variable	(1)		(2)		Normalized difference (1)-(2)	F-test for joint orthogonality
	Face-to-face N	Mean/SE	List sensitive group N	Mean/SE		
Wife can read and write	1412	0.159 [0.010]	1405	0.132 [0.009]	0.08	0.04**
Number of children	1412	4.674 [0.067]	1405	4.722 [0.065]	-0.02	0.60
Age	1412	35.118 [0.298]	1405	35.011 [0.288]	0.01	0.80
In polygamous marriage	1412	0.414 [0.013]	1405	0.396 [0.013]	0.04	0.35
In paid work last week	1412	0.004 [0.002]	1405	0.010 [0.003]	-0.07	0.07*
Household experienced accidents/disasters past year	1412	3.000 [0.030]	1405	3.070 [0.029]	-0.06	0.09*
Current value of household assets (Naira)	1412	22689.122 [846.079]	1405	21154.032 [755.367]	0.05	0.18
Husband's age when first married	1410	22.994 [0.159]	1403	24.243 [0.728]	-0.06	0.09*
Number of gender equitable attitudes held	1412	2.505 [0.016]	1405	2.528 [0.016]	-0.04	0.30
Locus of control	1412	10.794 [0.077]	1405	10.631 [0.078]	0.06	0.14

Notes: The value displayed for F tests are F statistics for the differences in means across groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure A3: Nigeria- relationships between IPV and respondent characteristics across survey methods

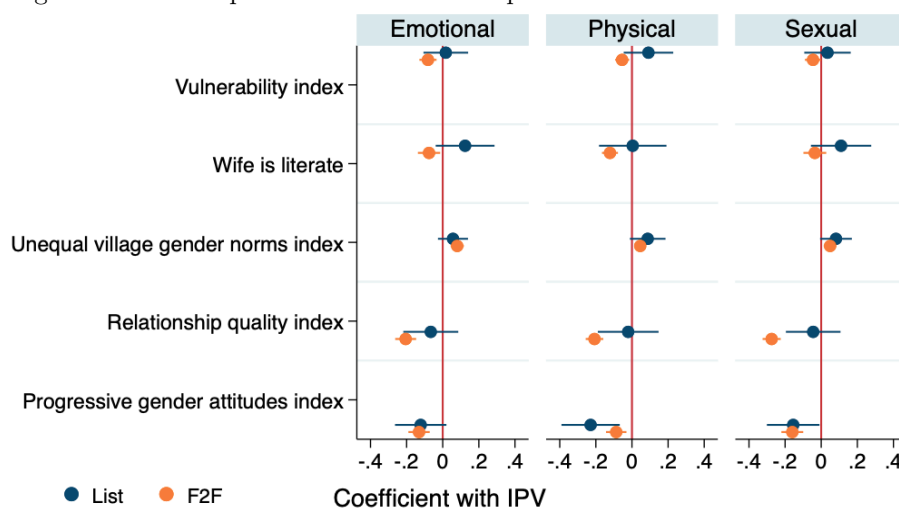


Table A9: Rwanda: Difference in mean physical IPV prevalence across methods and by respondent characteristics

Variable	Below median			Above median		
	(1) List	(2) F2F	(3) ACASI	(4) List	(5) F2F	(6) ACASI
Vulnerability index	0.12	0.08	0.10	0.28	0.10	0.11
Unequal village gender norms index	0.09	0.09	0.08	0.31	0.10	0.12
Relationship quality index	0.20	0.14	0.13	0.21	0.05	0.07
Progressive gender attitudes index	0.21	0.10	0.12	0.21	0.08	0.09
Gender engagement index	0.34	0.12	0.11	0.15	0.08	0.10
Wife's years of education	0.20	0.10	0.13	0.21	0.09	0.08
Wife's age	0.19	0.08	0.10	0.22	0.11	0.11
<i>Indicator variables</i>	Not legally married			Legally married		
Legally married	0.23	0.09	0.12	0.18	0.10	0.09
	Wife not in paid work			Wife in paid work		
Wife in paid work last week	0.27	0.09	0.09	0.11	0.09	0.12

Note: This table shows the difference in IPV prevalence by a number of respondent characteristics for the physical IPV question. The left column lists the respondent characteristics correlated with IPV. Columns 1-6 show the mean IPV prevalence by respondent characteristic (above or below median for continuous variables, or by binary characteristics). Columns 1 and 4 were estimated using a regression on the number of list items reported for the restricted sample of those above and below median.

Figure A4: Rwanda- Above and below median: difference in mean physical IPV prevalence across methods by respondent characteristics

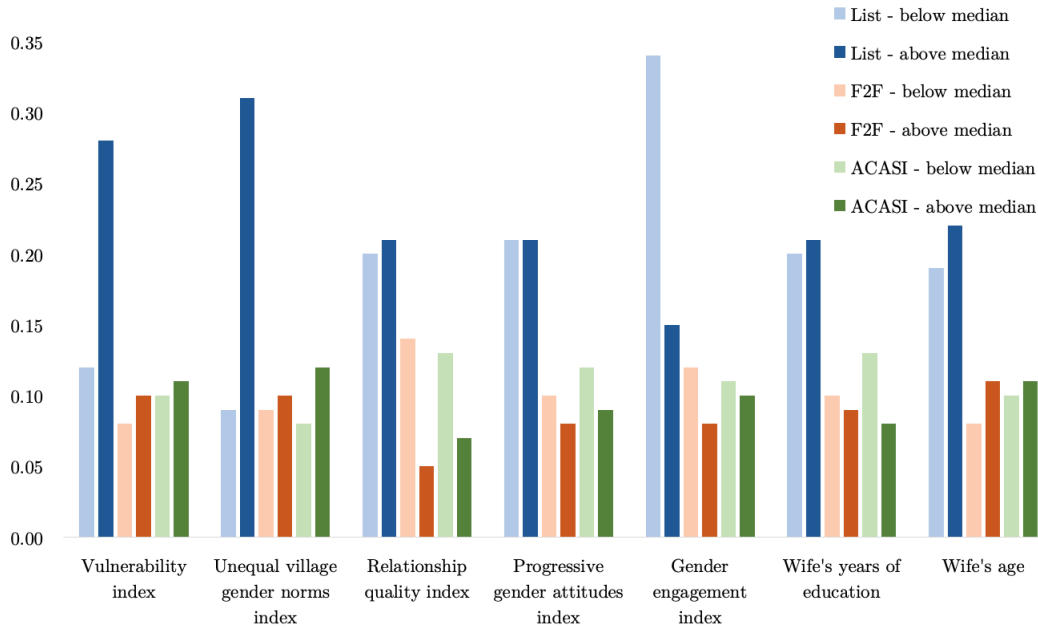


Table A10: Nigeria: Difference in mean IPV prevalence across methods and by respondent characteristics

Variable	Emotional IPV				Physical IPV				Sexual IPV			
	Below median		Above median		Below median		Above median		Below median		Above median	
	(1) List	(2) F2F	(3) List	(4) F2F	(5) List	(6) F2F	(7) List	(8) F2F	(9) List	(10) F2F	(11) List	(12) F2F
Vulnerability index	0.43	0.33	0.36	0.26	0.27	0.20	0.26	0.18	0.27	0.30	0.25	0.26
Unequal village gender norms index	0.40	0.24	0.39	0.36	0.24	0.15	0.29	0.23	0.24	0.27	0.28	0.30
Relationship quality index	0.44	0.39	0.36	0.22	0.29	0.27	0.24	0.13	0.28	0.39	0.25	0.19
Progressive gender attitudes index	0.41	0.35	0.38	0.25	0.30	0.24	0.23	0.14	0.28	0.35	0.24	0.21
Wife's age	0.36	0.27	0.43	0.32	0.21	0.18	0.31	0.21	0.23	0.28	0.29	0.27
Indicator variables	Wife illiterate	Wife illiterate	Wife literate	Wife literate	Wife illiterate	Wife illiterate	Wife literate	Wife literate	Wife illiterate	Wife illiterate	Wife literate	Wife literate
Wife reads and writes	0.38	0.31	0.48	0.24	0.27	0.21	0.24	0.09	0.25	0.29	0.33	0.25
Confidence in speaking up in village	Least confident	Least confident	Most confident	Most confident	Least confident	Least confident	Most confident	Most confident	Least confident	Least confident	Most confident	Most confident
Marriage type	0.41	0.34	0.38	0.21	0.25	0.21	0.29	0.15	0.27	0.29	0.25	0.27
Wife brought assets to marriage	Polygamous	Polygamous	Monogamous	Monogamous	Polygamous	Polygamous	Monogamous	Monogamous	Polygamous	Polygamous	Monogamous	Monogamous
	0.36	0.32	0.43	0.28	0.17	0.21	0.33	0.18	0.19	0.31	0.31	0.26
	No assets to marriage	No assets to marriage	Brought assets	Brought assets	No assets to marriage	No assets to marriage	Brought assets	Brought assets	No assets to marriage	No assets to marriage	Brought assets	Brought assets
	0.42	0.40	0.38	0.22	0.25	0.27	0.27	0.14	0.28	0.33	0.25	0.24

Note: This table shows the difference in IPV prevalence by a number of respondent characteristics, for each of the 3 IPV questions asked in Nigeria. The left column lists the variables correlated with IPV along the three types of violence. Columns 1-12 show the mean IPV prevalence by respondent characteristics (above or below median for continuous variables, or by binary characteristics). Coefficients in columns 1, 3, 5, 7, 9, and 11 were estimated using a regression on the number of list items reported.

Figure A5: Nigeria- difference in mean IPV prevalence by respondent characteristics: above and below median (left) and indicator variables (right)

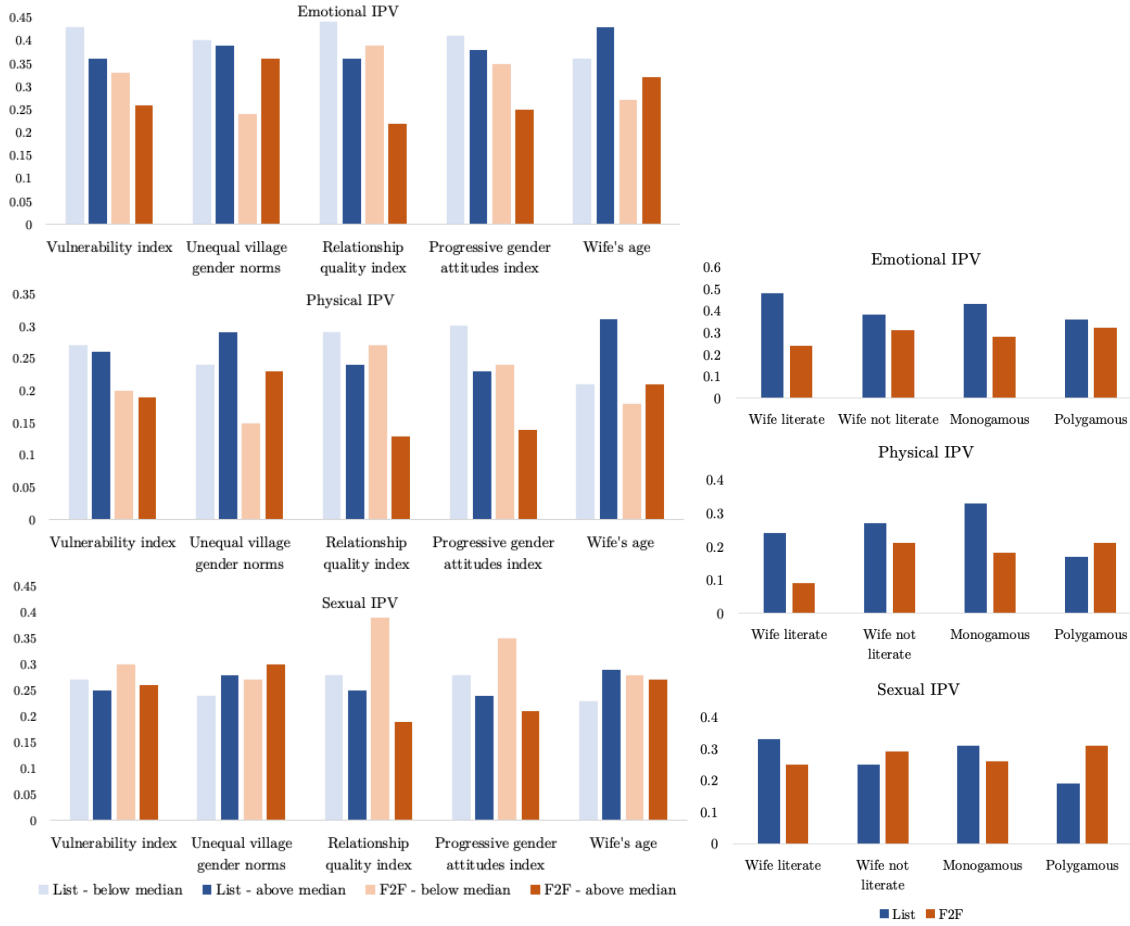


Table A11: Rwanda- No design effects test and proportions responding with each option

Response	Women (% of respondents)				Men (% of respondents)			
	Q1		Q2		Q1		Q2	
	Non-sensitive	Sensitive	Non-sensitive	Sensitive	Non-sensitive	Sensitive	Non-sensitive	Sensitive
0	3.67	6.76	13.69	19.84	8.97	8.55	11.24	16.88
1	28.03	18.9	60.22	43.81	39.22	21.32	46.14	32.5
2	53.05	43.99	22.26	28.78	44.57	40.82	34.74	38.65
3	15.26	27.84	3.83	6.65	7.24	24.97	7.89	11.29
4		2.52		0.92		4.33		0.68
No design effect p-value:		0.0012		0.0001		1		0.0001

Note: Table shows Bonferroni-corrected minimum P-values from no design effects test. Null hypothesis is of no design effect. Computed using List package on R (Blair & Imai, 2011)

Table A12: Nigeria- No design effects test and proportions responding with each option

Response	Nigeria (% of respondents)					
	Q1		Q2		Q3	
	Non-sensitive	Sensitive	Non-sensitive	Sensitive	Non-sensitive	Sensitive
0	13.81	8.9	12.54	8.61	11.47	11.25
1	59.99	40.64	47.95	38.01	50.14	38.72
2	19.62	37.15	24.86	35.59	32.93	33.67
3	6.59	9.54	14.66	12.38	5.45	13.17
4		3.77		5.41		3.2
No design effect p-value:		1		1		1

Note: Table shows Bonferroni-corrected minimum P-values from no design effects test. Null hypothesis is of no design effect. Computed using List package on R (Blair & Imai, 2011)

Table A13: Rwanda women - robustness checks with controls and enumerator fixed effects

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Q1: Past year emotional IPV			Q2: Lifetime non-partner sexual violence		
List method prevalence	0.206*** (0.035)	0.203*** (0.033)	0.203*** (0.035)	0.088*** (0.034)	0.102*** (0.031)	0.084** (0.033)
Observations	2,728	2,715	2,728	2,727	2,714	2,727
Controls		x			x	
Enumerator fixed effects			x			x

Notes: Controls selected using PDS lasso (Belloni et al., 2014). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A14: Rwanda men - robustness checks with controls and enumerator fixed effects

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Q1: Limited wife's family contact			Q2:Threatened to hurt wife/someone close		
List method prevalence	0.451*** (0.038)	0.440*** (0.034)	0.442*** (0.038)	0.071** (0.036)	0.065* (0.034)	0.067* (0.037)
Observations	2,728	2,721	2,728	2,728	2,721	2,728
Controls		x			x	
Enumerator fixed effects			x			x

Notes: Controls selected using PDS lasso (Belloni et al., 2014). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A15: Nigeria women - robustness checks with controls and enumerator fixed effects

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Q1: Past year emotional IPV			Q2: Past year physical IPV			Q3: Past year sexual IPV		
List method prevalence	0.397*** (0.032)	0.390*** (0.031)	0.415*** (0.026)	0.263*** (0.035)	0.258*** (0.035)	0.271*** (0.027)	0.260*** (0.032)	0.258*** (0.032)	0.275*** (0.025)
Observations	2,817	2,800	2,817	2,817	2,800	2,817	2,817	2,800	2,817
Controls		x			x			x	
Enumerator fixed effects			x			x			x

Notes: Controls selected using PDS lasso (Belloni et al., 2014). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A16: Refusal rates by question and method

	Face-to-face		ACASI		List	
	No. respondents	%	No. respondents	%	No. respondents	%
Rwanda						
Women Q1	0	0	23	0.84	0	0
Women Q2	1	0.04	23	0.84	1	0.04
Men Q1	0	0	66	2.41	0	0
Men Q2	0	0	16	0.59	0	0
	Face-to-face		List			
	No. respondents	%	No. respondents	%		
Nigeria						
Q1	5	0.18	0	0		
Q2	5	0.18	0	0		
Q3	18	0.64	0	0		

B Supplementary discussion on IPV measurement

The evidence presented in this paper regarding the substantial and systematic misreporting on IPV in Rwanda and Nigeria suggests several issues for consideration by those planning to measure IPV and other sensitive topics.

What trade-offs should social scientists and policy makers consider when deciding what method to use when measuring self-reported IPV? The face-to-face method is the simplest approach. Unlike the list and ACASI methods, it asks questions in the same format as most other survey questions and thus does not require a lengthy period of explanation by the enumerator. Nor does it require electronic devices or headphones, nor some minimum cognitive ability for respondents to understand more complex question formats. However, as this experiment shows, the face-to-face method likely results in the most underreporting.

Despite expectations, women’s self-reported IPV experience did not significantly differ between ACASI and face-to-face methods. This suggests that for researchers whose objective is to reduce misreporting bias in IPV measurement, ACASI may not be worth the additional cost (though there may be ethical, logistical, or other reasons why ACASI may be preferred).

Is the list experiment then the optimal IPV data collection method? This paper suggests that of the methods tested here, the list experiment appears to most minimize underreporting. However, there are a number of challenges associated with its design, administration, and analysis that may make it inappropriate for some purposes or in some contexts. For example, the list method is more time-consuming and cognitively demanding relative to other methods.

Other challenges with the list experiment have been well documented by Blair and Imai (2012); Blair, Imai and Lyall (2014); Blair, Coppock and Moor (2018) and Chuang et al. (2018). However, one particular issue should be highlighted: the list experiment typically has greater variance than direct questioning methods, but also likely lower misreporting bias. Blair, Coppock and Moor (2018) suggest that if the expected misreporting bias is between 5 and 10 percentage points, direct questions are preferred if sample sizes are less than 3,000. In this paper, we found that the bias for IPV measurement was larger than 10 percentage points in some cases, suggesting IPV measurement may be an area where the method’s use is justified. However, for those using list experiments to estimate treatment effects, adequately powered studies likely require even larger sample sizes given that list method groups get split again by treatment arms. At the very least, researchers and policy makers can use these results to help inform their survey method choice.³⁴

It is worth noting that for researchers measuring IPV using the multi-question DHS module, an equivalent set of list experiment questions would not allow for the construction of an ‘any violence’ indicator variable. Traditionally, most studies aggregate women’s direct yes/no responses in each category

³⁴In addition to statistical power concerns with the list experiment, there are the well-documented challenges with designing appropriate non-sensitive list questions to prevent ‘floor’ and ‘ceiling effects’ so that respondents’ answers are disguised. There are also debates about the optimal types of non-sensitive items, with some evidence suggesting that the non-sensitive items should be somewhat sensitive so that the question of interest is not salient, with associated concerns about the sensitivity of list method estimates to differences in types of non-sensitive questions. It is also possible that ordering of the sensitive question within the list matters, as well as the number of sets of list questions asked. For example, in this study, given each respondent was assigned to either the sensitive or non-sensitive list group, there may be reporting effects for the subsequent lists a respondent was read, with respondents perhaps becoming increasingly suspicious of the method, though in Nigeria, list method prevalence did not substantially decline over the 3 questions on emotional, physical and sexual violence. See Coutts and Jann (2011); Glynn (2013); Blair and Imai (2012); Blair, Coppock and Moor (2018) and Chuang et al. (2018) for a fuller discussion on list experiment trade-offs and design.

of emotional, physical, and sexual violence to create summary indicator variables. This cannot be done with the list experiment as respondents' direct answers are unknown. Replacing the traditional aggregated approach, which has been recommended because it gives women multiple chances to disclose violence, with list estimates may replace one form of bias (underreporting on individual questions) with different bias (from not being able to create a summary indicator reflecting the extent of violence experienced by each respondent). Further research could seek to establish how these biases relate and their respective impacts on prevalence and correlation estimates, and whether there are methods to simultaneously minimize both.

Finally, while the list method may minimize misreporting bias in correlations and treatment effect estimates caused by fear or shame, it cannot mitigate bias in treatment effects from interventions that change women's understanding or interpretation of terms used in IPV outcome questions such as 'humiliation', 'forced sex', or 'threatening', for example. Yet both of these sources of reporting bias- 'liberating' (from shame and fear) and 'learning' (from a changed interpretation of question wording or broadening awareness)- may confound treatment effect estimates of interventions' impact on IPV. The list method is only likely to minimize the former source of bias while leaving the latter unchanged.