NBER WORKING PAPER SERIES

MEASUREMENT ERROR IN SELF-REPORTED DATA ON STIGMATIZED BEHAVIOR: EXPERIMENTAL EVIDENCE ON INTIMATE PARTNER VIOLENCE

Jorge M. Agüero
Ursula T. Aldana
Erica M. Field
Veronica Frisancho
Ignacio Rodriguez Hurtado
Javier Romero

Working Paper 35891
http://www.nber.org/papers/w35891

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2026

This paper is dedicated to the memory of Jorge Agüero, a cherished colleague whose foundational work and intellectual contributions made this research possible. Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health (under Award Number R01HD101581), the Innovations for Poverty Action Intimate Partner Violence Initiative, the Inter-American Development Bank, the J-PAL Crime and Violence Initiative, the J-PAL Government Partnership Initiative, the SAMRC Sexual Violence Research Initiative, the World Bank's Latin America and the Caribbean Gender Innovation Lab, and the World Bank's Umbrella Facility for Gender Equality (UFGE). The UFGE has received generous contributions from Australia, Canada, Denmark, Finland, Germany, Iceland, Ireland, the Netherlands, Norway, Spain, Sweden, Switzerland, the United Kingdom, the United States, the Bill and Melinda Gates Foundation, and the Wellspring Philanthropic Fund. Gustavo Acosta, Daniel Hurtado, Gshan Irigoin, Glenn Ortiz, Cindy Rojas, Juan Pablo Rossi, and Hugo Salas, provided excellent research assistance at various stages of this project. The findings, interpretations, and conclusions expressed in this work are entirely those of the authors and should not be attributed in any manner to the National Institutes of Health, the World Bank, the World Bank's Board of Executive Directors, or the governments they represent. All errors are our own. Duke Institutional Review Board number: 2018-0262. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Jorge M. Agüero, Ursula T. Aldana, Erica M. Field, Veronica Frisancho, Ignacio Rodriguez Hurtado, and Javier Romero. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Measurement Error in Self-Reported Data on Stigmatized Behavior: Experimental Evidence on Intimate Partner Violence

Jorge M. Agüero, Ursula T. Aldana, Erica M. Field, Veronica Frisancho, Ignacio Rodriguez Hurtado, and Javier Romero

NBER Working Paper No. 35891
October 2026
JEL No. C81, C83, J12, J16, O12

ABSTRACT

Error in the measurement of key development outcomes affects our understanding of socioeconomic phenomena and their policy implications. This paper focuses on measuring intimate partner violence (IPV), a public health issue affecting women globally. We introduce a novel, privacy-enhancing data collection method using a ballot, comparing it to standard survey techniques in which responses are not kept private from enumerators, and list experiments, an approach ensuring complete anonymity. Our findings show that improved privacy yields higher and more accurate IPV reports than both, especially among women with low education. By linking our data with two natural experiments, we demonstrate that the non-random nature of underreporting with standard methods can lead to biased estimates of IPV determinants.

Jorge M. Agüero
University of Connecticut

Ursula T. Aldana
Institute of Peruvian Studies
ualdana@iep.org.pe

Erica M. Field
Duke University Department
of Economics and CEPR
and also NBER
emf23@duke.edu

Veronica Frisancho
Inter-American Development Bank
vfrisancho@iadb.org

Ignacio Rodriguez Hurtado
Charles River Associates
ignaciorodriguezhurtado91@gmail.com

Javier Romero
The World Bank
Latin America and the Caribbean
Gender Innovation Lab (LACGIL)
and Lima School of Economics
at Universidad de Piura
javierromero@worldbank.org

An online appendix is available at http://www.nber.org/data-appendix/w35891
A randomized controlled trials registry entry is available at AEARCTR-0002629

1 Introduction

There is growing concern about the quality of survey data, which some scholars believe is declining, particularly with respect to sensitive topics [Meyer et al., 2015]. Measurement error driven by reluctance to self-report on sensitive topics not only limits our ability to conduct scientific research in these domains but also has the potential to lead to misallocation of government spending on social programs that rely on survey data to correctly identify target populations.

In this paper, we study measurement error caused by individuals' reluctance to accurately report their experiences with Intimate Partner Violence (IPV), and propose an alternative cost-effective way to collect less erroneous data on the issue by increasing respondent privacy. We also document the consequences for empirical research on IPV by evaluating the nature of bias that results from standard survey methods.

IPV is an extremely prevalent global phenomenon that has been identified as a major challenge to global public health by the World Health Organization. Measurement error is a concern for survey data on IPV as victims' sense of stigma or embarrassment may lead to non-random measurement error through under-reporting. Moreover, as many scholars have argued, IPV is prone to particularly high rates of misreporting [DeKeseredy and Schwartz, 1998; Aizer, 2010, 2011; Kishor, 2015]. First, IPV tends to take place behind closed doors, so it is widely unobservable to the larger community, such that victims are more likely to be held responsible for disclosure. Second, the aggressor is a person close to the victim - a factor that, when combined with the previous one, raises the potential emotional, financial or physical costs of disclosure for the victim. In addition, the importance of measuring IPV through survey data is greater than it is for other types of crime in light of the absence of reliable administrative records on its prevalence, a consequence of victims' low rates of police reporting.[1]

Our study conducts a survey experiment among 5,562 respondents in rural Peru that allows us to compare the performance of two alternative methods designed to improve the accuracy of self-reports on sensitive topics. In particular, we collect self-reported data on physical and sexual IPV using three different survey methods randomly assigned across households within villages: (1) face-to-face oral interview, that closely adapts the widely employed Domestic Violence Module from the Demographic and Health Surveys (DHS); (2) the list experiment method, a more costly survey instrument showcased in the literature as a means of guaranteeing full anonymity; and (3) the ballot approach, a novel method that ensures respondent privacy, but does not guarantee anonymity, and is significantly more cost-effective and data-rich than list experiments.

All women were administered a series of IPV questions modeled after the Conflict Tactic Scale (CTS) [Straus et al., 1996].[2] The CTS asks about specific violent episodes (e.g., "Did your partner ever push you?"), a procedure that has been shown to significantly improve reporting accuracy compared to more general questions on IPV in which the definition of violence is more subjective and thus open to respondent interpretation (e.g., "Have you experienced violence?"). In one of the treatment arms, the survey was conducted following the DHS protocol, including face-to-face interviews conducted in private in which the enumerator recorded women's answers as they responded aloud. Given its characteristics,

[1] Palermo et al. [2014] show that only a very small proportion of victims report IPV cases to the authorities and that there is a strong self-selection into reporting.

[2] The CTS has been applied in over 122 surveys covering 61 developing countries, including Peru, as part of the DHS project.

we label this arm "Audible Response". Although widely considered the gold standard for measuring IPV, this protocol may not fully address respondents' psychic cost of discussing the issue out loud and with a stranger, which could lead to downward-biased reporting of incidence. Psychic costs of reporting on IPV may stem from multiple sources, including fear of repercussions from perpetrators if respondents fear their interviews can be overheard, social desirability bias heightened by shame or embarrassment of admitting to a stigmatized behavior, or self-denial/reluctance to admit to oneself something that feels beyond one's control [Kishor, 2015; Agüero and Frisancho, 2022; Bursztyn et al., forthcoming].

To test for potential reporting bias, we implement two alternative survey modules that provide respondents with additional levels of privacy. In the second arm, we offer respondents full anonymity through an indirect questioning technique called a *list experiment*, a survey method that is increasingly used in the field to measure the incidence of experiences that are likely to involve stigma or shame and hence thought to be widely under reported, including use of drugs and alcohol, violence, crime, anti-gay sentiment, and stigmatized behaviors such as female genital cutting or self-identification as LGBTQ+ [e.g., Karlan and Zinman, 2012; Druckman et al., 2015; De Cao and Lutz, 2018; Coffman et al., 2017; Gibson et al., 2018; Cullen, 2023]. List experiments guarantee full anonymity as their indirect nature allows neither the surveyor nor the researcher to identify prevalence rates at the individual level.[3] We label this second arm "Anonymous Response".

However, the greater levels of privacy offered by the list experiment come at several costs. The first major cost is the lower precision of prevalence estimates [Blair et al., 2020]. The second important cost of list experiments is that they only deliver average prevalence rates rather than individual-level data. This has two major disadvantages critical to research on IPV: it complicates the analysis of correlations between individual characteristics and episodes of violence, and it fails to produce total measures of violence (e.g., experiencing any type of violence) since one cannot link responses for each act of violence across the same individual.[4]

The third major cost of implementing list experiments is the difficulty of explaining to respondents how to answer the questions, since the instructions are not simple and respondents tend to lack experience with this type of exercise. Chuang et al. [2021], Blair et al. [2019] and Ahlquist [2018] provide evidence of important levels of measurement error with data from list experiments in different contexts. In order to be effective, enumerators and respondents need to be trained in the art of adding up items and reporting the sum, and made to understand that reporting a sum across items preserves anonymity, which may not be obvious at first blush. Any remaining confusion over the technique or distrust of the method increases random measurement error and/or biases downward reporting of IPV, which could easily undo any benefits of providing anonymity. In this vein, Gilligan et al. [2025] interpret the lower rates they observe in list experiment data (compared to the standard survey method) as a possible consequence of lack of comprehension.

Our third approach to eliciting IPV circumvents these limitations while still providing full privacy, though not anonymity. In particular, in the third arm we introduce a novel form of data collection via hidden reporting through a ballot, borrowing from voting-style techniques in the public health literature [Barr et al., 2017; Gregson et al., 2004; Phillips et al., 2013; Singh et al., 2010]. Respondents in this arm were given a sheet of paper with

[3] See Blair and Imai [2012] for a review of the methodology and applications.

[4] Even though there are methods for examining the correlation between individual characteristics and the sensitive item when using the list experiment (see Blair and Imai [2012]), this analysis is more direct and efficient when individual-level information is available.

a unique individual identifier and were instructed to mark whether specific IPV items had occurred after verbal prompting of each. Subsequently, they deposited the paper into a sealed box, a practice that mimics voting in an election.

We label this arm the "Private Response" since this method serves as a middle ground between the "Audible Response" and the "Anonymous Response". It offers full privacy but not anonymity. Moreover, the method has the critical advantage of being relatively simple to implement in the field because instructions are easy to follow and respondents in most settings are accustomed to this type of exercise through voting procedures in government elections, thereby reducing measurement error due to respondent confusion relative to the list experiment approach.

Our results reveal that the "Private Response" method yields IPV rates that are systematically higher than those generated by the "Audible Response" method. Rates are uniformly higher across item-specific questions, with statistically significant differences for six of seven IPV statements. Overall, offering privacy through the ballot increases the rate of reporting any type of IPV by 17%. Moreover, the "Private Response" approach yields statistically comparable rates for some item-specific violence relative to the "Anonymous Response".[5] In other cases, the "Anonymous Response" rates are *lower* than "Audible Response" estimates. This suggests that respondent confusion introduced a non-trivial, non-random measurement error that encouraged underreporting and thus counteracted the positive effect of anonymity. In contrast, the ballot method appears to generate more consistently accurate data on IPV prevalence than both alternative methods by reducing under-reporting without generating respondent confusion. Additionally, it overcomes two of the main limitations of list experiments: low statistical precision and lack of individual-level data. Evidence from an additional survey conducted in 2025 and matched to our original data shows that survey miscomprehension is not driving the results.

More generally, the comparison across methods provides novel evidence that greater privacy alone can reduce measurement error. This is important because, if privacy matters more than anonymity, a range of survey techniques beyond ballot methods could have a significant impact on reporting rates, including voice-assisted survey data collection on tablets as well as simpler protocols such as allowing respondents to mark rather than speak responses to survey questions.

In the final section of the paper, we explore how using "Private Response" in place of "Audible Response" data on IPV leads to qualitatively distinct inference in analyses of IPV determinants. In particular, we make use of two different natural experiments that have been studied in the literature, which shift potential drivers of IPV in our sample. First, we study how IPV prevalence changes in response to arguably exogenous variation in household resource levels that arises from the district-level timing of payments from *Juntos*, a national cash transfer program. We show that, under "Private Response" conditions, there is a clear and significant positive relationship between household resource levels and violence, while a similar analysis using responses to the "Audible Response" questions yields no significant pattern.

Second, we make use of a natural experiment, also utilized by Weitzman [2018], in which a cohort of students was pushed into one additional year of schooling due to a change in mandatory schooling laws. We show that the IV estimate of the causal effect of this reform on IPV using "Audible Response" data indicates that education reduces violence,

[5]Since item-specific responses cannot be aggregated for each individual, the list experiment method does not generate data on "any IPV" .

whereas relying on data generated by "Private Response" masks this association. Both results underscore how measurement error can generate biased estimates not only of IPV levels but also confound causal inference of IPV determinants.

A recent strand of the literature has used alternative methods to elicit truthful reporting of IPV. A few studies have relied on list experiments to elicit IPV prevalence and compare it against "Audible Response" interviews with mixed results [Cullen, 2023; Agüero and Frisancho, 2022; Gilligan et al., 2025]. In turn, others have questioned the ability of list experiments to produce valid estimates, as explained above. Alternative data collection methods such as Audio Computer-Assisted Self-Interview (ACASI) have produced significantly higher IPV rates compared to "Audible Response" reporting [Peterman et al., 2024; von Russdorf et al., 2024]. However, the applicability of ACASI techniques has also been called into question due to survey miscomprehension [Park et al., 2024]. A recent study in Pakistan by Iqbal et al. [2025] developed and piloted an adaptation of ACASI that enhances respondent understanding, underscoring the importance of minimizing respondent confusion when employing alternative methods.

This paper makes two important contributions to the economics literature. First, it develops and tests the superiority of a ballot, or "Private Response," method for survey-based IPV measurement, and shows that it performs better than existing approaches for measuring IPV in surveys.[6] The method, which resembles a voting exercise using paper ballots, has roots in public opinion surveying [Benson, 1941; Bishop and Fisher, 1995] and has been used in public health research to elicit information on sexual and reproductive behaviors.[7] Despite their potential to improve accuracy and truthful reporting, voting-style and ballot-based techniques remain rare in economics. Exceptions include Olken [2007], who uses a sealed-envelope reporting technique as part of an intervention to reduce corruption in Indonesia, and Crépon et al. [forthcoming], who use a fully anonymous ballot method in Egypt to study irregular migration intentions.[8] An important distinction between these approaches and ours, other than the stigmatized behavior being measured, is that our approach does not guarantee anonymity, allowing us to rule out the role of anonymity in bolstering reporting of IPV and potentially other stigmatized behaviors, while also enabling researchers to directly tag individual respondents to ballot responses for analysis.

Second, to the best of our knowledge, this is the first paper to measure the bias that can arise from the evaluation of policy interventions when the outcome variable is subject to misreporting. Our comparison of the results obtained using "Private Response" and "Audible Response" data provides insights into the nature of IPV underreporting that occurs when using traditional approaches, which has implications for empirical studies of the issue more generally.

[6] While Cullen [2023] compares ACASI with both list experiments and "Audible Response" methods, it does so only for one of several physical IPV items in the DHS. Our study investigates the comprehensive list of IPV items widely used to capture total rates of IPV.

[7] For example, see Barr et al. [2017]; Gregson et al. [2004]; Phillips et al. [2013]; Singh et al. [2010]; Cagney et al. [2025]; Chowdhury et al. [2012]; Lépine et al. [2020].

[8] The study by Crépon et al. [forthcoming], which is the most similar to our methodology, was conducted after our experiment, in 2022, and analysis conducted concurrently to this study.

2 Experimental Design

2.1 Sample and Outcome Measures

In comparing survey methods, our primary outcomes of interest are the incidence of any physical and sexual violence, in line with much of the existing literature on IPV. We exclude non-physical forms of IPV—such as emotional, psychological, financial, and cyber violence—because of greater definitional clarity across all items included in standard measurement of physical and sexual violence. Items included in the measurement of emotional violence, for instance, typically include more abstract concepts and subjective experiences such as humiliation, threats, or control, which can be difficult for individuals to understand clearly in survey instruments. Questions about emotional abuse may thus yield less consistent responses, posing challenges to measurement validity and comparison.

To study measurement error in these outcomes, we collected information from 5,562 women living in rural and peri-urban Peru in 2018. As shown in Online Appendix Table A1, our study sample comprises 12 rural districts distributed across 9 *Departamentos*, which represent the highest official administrative divisions in Peru.[9] Within these districts, 260 villages were selected as part of an impact evaluation of a program implemented by the Peruvian Ministry of Women and Vulnerable Population (MIMP). Our data were collected as part of its baseline census. Sample districts belonged to the catchment area of 12 Women's Emergency Centers (CEMs in Spanish).[10] The sample frame consisted of all women aged 18-49 who had children and were currently cohabiting with a partner.[11]

Since our study sample is not nationally representative, it is informative to compare IPV prevalence in our "Audible Response" arm to the rates observed in the corresponding Peruvian DHS (2017). It is worth noting that respondents in our sample report experiencing substantially higher rates of physical and sexual violence relative to the 2017 DHS sample (0.39 versus 0.31) captured using the same survey method, which likely reflects a combination of time trends in IPV, which was trending upwards over this period, and selection criteria particular to our sample (e.g. regional balance). Thus it is important to bear in mind that the biases we document may be more pertinent to our sample.

2.2 Randomization and Survey Instrument

In each village, women were randomly assigned to one of three experimental arms that determined which IPV module they received as part of a larger household survey questionnaire. All questionnaires began with three identical modules designed to mimic the structure of the DHS questionnaire: demographics, women's health (unrelated to IPV), and intra-household decision-making. In module 4, subjects were randomly assigned to one of three methods of inquiring about IPV-related incidents. In all three arms, we followed DHS/WHO guidelines for IPV-related surveys.[12] We use a range of variables collected in the survey to check for

[9] *Departamentos* include Amazonas, Apurímac, Ayacucho, Cusco, Huancavelica, Huánuco, Junín, La Libertad, and San Martín.

[10] Selection of districts into the study was based on the number of villages per district and the number of grassroots organizations needed for the impact evaluation.

[11] We dropped 2.4% of the original study sample (134 women), as they belonged to villages that were too small to randomize all three arms of our survey experiment within each village. This allows us to conduct inference using the same set of villages across all treatment arms. Dropping these observations does not affect our results or the balance of the experiment.

[12] In addition to the regular procedures in place for human subject interviews, these guidelines include providing a referral to services available in the area and a strict protocol to secure women's privacy. The complete guidelines implemented in the study are available upon request.

balance across experimental arms. Online Appendix Table A2 confirms that the samples are balanced across experimental arms along the majority of demographic characteristics, and a joint test fails to reject the null hypotheses that the experimental groups have the same distribution of observables. Similarly, Online Appendix Table A3 compares the two list experiment subarms (lists with and without the IPV item) and shows they are also balanced.

2.2.1 Standard Survey Method

The first arm is the "Audible Response" arm and its questionnaire was identical to the DHS item-specific IPV module, asking respondents to report out loud whether they had experienced specific acts of physical and sexual violence. Consistent with the DHS, our survey questions referred to *ever* experiencing violence with the most recent partner, a recall period that recent evidence suggests may generate substantial longitudinal inconsistencies in IPV self-reports [Abramsky et al., 2022]. This arm serves as a benchmark, allowing us to compare whether higher levels of privacy and anonymity increase the reported incidence of IPV.

2.2.2 List Experiment

Respondents assigned to the "Anonymous Response" arm were administered a list experiment for each of the IPV questions in place of the "Audible Response" questions. In this arm, subjects were presented with a list containing either four (group A) or five (group B) statements and asked to report the total number of statements that were true. As with all list experiments, while the researcher cannot observe whether any individual in this arm has experienced violence, they can infer a *rate* of violence within a group of respondents by comparing the difference in responses across groups A and B. For instance, if the difference in mean responses across groups is 0.4, this would imply that 40% of respondents in group B have experienced the IPV item that was added to their list.[13]

The list experiment approach offers respondents full privacy since they at no point reveal whether or not they have experienced a particular act of violence. They are also almost always provided with full *anonymity*, since the only way that an enumerator or anyone else who accesses the survey data would be able to deduce the individual incidence of IPV is if the respondent's answer was equal to the total number of list items. In building the list items, we minimized this possibility by including at least one pair of items with a high probability of being mutually exclusive. We also field tested the list to prevent ceiling effects in the lists.

While privacy and anonymity are presumed benefits of this method, it poses important drawbacks. First, since individual responses are not identifiable, this method only allows researchers to estimate a prevalence rate within a group rather than a specific value for each respondent.[14] Second, the protocol to implement the list experiment method, based on the number of true statements in a list, makes it difficult for respondents to trust and/or understand. It is not obvious that this method will be clear or trusting enough to really

[13] See Online Appendix Section 2 for more details on the list experiment protocol employed in this survey.

[14] To clarify this point, let S_{ij} be equal to one if woman i $(i = 1, ...N)$ experienced act of violence j $(j = 1, ...J)$. With standard survey questions, a researcher observes S_{ij} , as a binary variable, but it could be measured with error. With list experiments, it is not possible to observe S_{ij} but rather, only the rate of violence, defined by $S_j = (1/N) \sum_i S_{ij}$.

provide anonymity [Chuang et al., 2021]. This structure also increases the variance of the estimate, as we discuss with more detail in Section 3.1.

2.2.3 Ballot Method

In the third arm, the "Private Response" method, we implement an alternative method of collecting survey data on IPV that offers a higher degree of respondent privacy without providing full anonymity. To replicate the "Audible Response" method as closely as possible, enumerators read each IPV question out loud and immediately hand respondents a single ballot on which the respondent privately marks her answer and then folds and deposits it into the open slot of a sealed box. Each sheet contains the written IPV statement alongside a picture depicting the event. See Online Appendix Section 3 for more details on the ballot protocol.[15]

This method avoids the risk of someone in the household overhearing their report on experiencing IPV, but it also conveys to respondents that the enumerators themselves will not observe their answers, increasing respondent's perceived privacy.[16] Perhaps most importantly, ballot respondents do not have to face the psychic cost of speaking openly about an event that may evoke shame, stigma, or embarrassment. However, unlike entirely self-administered approaches, respondents are invited to ask clarifying questions to the enumerator after she poses the question out loud.

To the best of our knowledge, this is the first application of such an approach to collecting survey data on IPV. It is also the first time it has been tested directly alongside audible and anonymous response methods. The ballot approach has several advantages over the list experiment. First, it generates individual measures of each act of violence, as the paper ballot responses are later matched to individual questionnaires through a unique identifier.[17] Second, the ballot approach is significantly easier to implement than a list experiment since it does not involve the cognitive effort of summing across list items or potential confusion stemming from learning and executing an unfamiliar cognitive exercise. Third, the prevalence estimate obtained using the ballot method has a lower variance, as we discuss in more detail in the next section.

A comparable method that has recently been employed in the literature on measuring stigmatized behaviors is the ACASI approach. This methodology is similar to the ballot method in that respondents do not openly discuss IPV with an enumerator, thereby ensuring greater respondent privacy. There are, however, a number of key differences that could favor one method over the other, depending on the setting and target population. Perhaps most importantly, the ACASI approach relies on voice recorded questions played through headphones, whereas in the ballot method the enumerator reads the questions out loud to respondents. An advantage of headphones is eliminating risk that other household members learn what is being asked. However, requesting respondents to wear headphones may draw attention and raise distrust among other household members.[18] One *dis*advantage of ACASI

[15] Pictures were not used in the "Audible Response" and "Anonymous Response" arms as they did not require the participant to answer by marking on a piece of paper. It is unlikely that the results presented in the paper are driven by the addition of these images rather than the privacy condition of the ballot arm. The questions used are based on the Conflict Tactics Scale with very specific, behaviorally concrete questions ("Did your partner ever push you?"). These questions are not conceptually ambiguous and have been used and validated in several settings.

[16] We make this distinction because, although it is true that the enumerators will never process the data obtained through the ballot, we cannot rule out that respondents believe that the same enumerator will remove and read their ballot response post-interview.

[17] We preprinted respondent IDs in the corner of each paper ballot.

[18] Iqbal et al. [2025] try to protect the privacy of their respondents by adding some non-sensitive questions to the

over ballot is that respondents generally progress independently through the survey module as enumerators are not supposed to interact with them unless a pause is requested to ask for clarification, potentially leading the respondent to rush through the module without adequate reflection or understanding [Park et al., 2024].[19] In the "Private Response" arm, each question is read aloud by the interviewer for the respondent to mark her answer in a ballot, allowing enumerators to better detect if respondents are understanding each question. Another important disadvantage of ACASI over ballot is that it requires respondents to have digital skills, a requirement that can exacerbate existing measurement error due to digital literacy divides across age, socioeconomic status, and geographic location.

3 Results

3.1 Rates Of Violence By Survey Method

The results of our measurement experiment are presented in Table 1. Columns 1 and 2 show rates of overall physical and sexual violence as well as item-specific IPV as reported using the traditional or "Audible Response" approach and the "Private Response" approach. Column 3 shows prevalence rates obtained for each type of violence using the "Anonymous Response" method.

Overall, IPV prevalence rates obtained using face-to-face interviews are everywhere lower than those obtained under the ballot method. A comparison of columns 1 and 2 reveals that the ballot approach yields higher point estimates for each type of violence, and the difference is statistically significant for six of seven IPV types. The private method also results in a 6 percentage point (or 17%) higher rate of any IPV and higher reported numbers of IPV types, both significant at the 1% level. In other words, privacy without anonymity strongly increases respondents' willingness to report IPV in household surveys. In general, the differences appear to be larger for more severe forms of violence. For instance, women in the "Private Response" arm are more than twice as likely as those in the "Audible Response" arm to report that their partner threatened them with a weapon and 50% more likely to report that they were burned. This pattern is consistent both with greater shame and greater fear driving under-reporting when implementing audible response methods.

A second finding from Table 1 is that the privacy offered through the ballot approach also outperforms the full anonymity provided by list experiments in measuring IPV. The list experiment method generates higher prevalence rates than standard interview questions (columns 1 versus 3) on only three out of seven questions, and none of these differences is statistically significant. Meanwhile, for four acts of violence, prevalence rates seem *lower* when measured with the list experiment compared to standard interview techniques.

start of the ACASI module to help diffuse distrust among other household members who at times demanded to hear the questions being asked.

[19] Cullen [2023] and Iqbal et al. [2025] show that using graphical aids can facilitate comprehension even among illiterate populations.

Table 1. Prevalence Rates By Survey Method

Variable	(1) Audible	(2) Private	(3) Anonymous	P-value (1)-(2)	P-value (1)-(3)
Any IPV	0.393	0.459	.	0.000	.
	(0.015)	(0.016)	(.)		
Push	0.285	0.346	0.249	0.000	0.145
	(0.013)	(0.014)	(0.024)		
Slap	0.232	0.239	0.188	0.613	0.132
	(0.012)	(0.012)	(0.027)		
Punch	0.214	0.242	0.249	0.051	0.104
	(0.012)	(0.012)	(0.021)		
Kick	0.157	0.177	0.180	0.090	0.424
	(0.010)	(0.011)	(0.028)		
Burn	0.042	0.062	0.068	0.010	0.122
	(0.005)	(0.006)	(0.017)		
Weapon	0.016	0.036	0.011	0.000	0.813
	(0.003)	(0.005)	(0.020)		
Forced Sex	0.101	0.131	0.093	0.002	0.665
	(0.008)	(0.008)	(0.018)		
N IPV statements	1.047	1.233	.	0.001	.
	(0.049)	(0.056)	(.)		
Observations	1853	1812	1897		
Clusters	260	260	260		

Notes: Clustered standard errors at the village level in parenthesis. Audible Response indicates traditional survey questions. Private Response indicates our experimental ballot method. Anonymous Response indicates the list experiment. Estimates of prevalence rates under list experiments are obtained from a regression of the indirect answer on the treatment dummy. Estimates of prevalence rates under the other two methods are obtained from a regression of the reported IPV on a constant. The set of IPV questions are described in Table A4 in the Online Appendix.

Table 1 also illustrates one of the main shortcomings of the list experiment method: its higher variance. The standard errors of each act of violence are much higher under the list experiment than under the other two methods despite the list experiment having a slightly larger sample size. The reason for the higher variance is that the list experiment method involves subtracting the mean of group A from the mean of group B, which implies that two variances are being summed and each is higher due to its smaller sample size (only half of the sample). Moreover, the variance depends not only on the sensitive question but also on the additional variation of the neutral statements. Online Appendix Section 4 details the role these two factors play in increasing the standard error of the list experiment.

We have thus far assumed that higher reporting of a stigmatized and sensitive topic like IPV reflects a reduction in underreporting, and therefore lower measurement error relative to the alternative techniques. To support this assumption, we investigate whether survey miscomprehension drives the difference between the ballot method and the standard technique.

To determine whether misunderstanding of the ballot protocol is driving our findings, we rerun the previous analysis excluding *inconsistent* respondents (see Online Appendix

Section 5 for details). In particular, we assess participants' understanding of the ballot protocol through three non-sensitive questions administered via ballot in a subsequent survey conducted in 2024/2025.[20] We categorize a respondent as inconsistent if any of their ballot reports on non-sensitive questions are incorrect; in general, the proportion of inconsistencies for non-sensitive questions is low, fluctuating between 2.2% and 4.8% across the three statements.[21] The results show that the differences between the "Private Response" and the "Audible Response" reports of physical and sexual IPV remain unaltered after excluding inconsistent respondents. Additionally, we verify that our results are robust to a range of controls and fixed effects, as shown in Online Appendix Table A5.

3.2 Who Are The Marginal Respondents?

We next investigate the characteristics of participants that are encouraged to report IPV when offered the "Private Response" method. To conduct this analysis, we borrow from traditional instrumental variables (IV) models, which have well defined tools to identify complier characteristics [Angrist et al., 1996; Marbach and Hangartner, 2020]. We use the ballot indicator as the instrument and reported IPV as the endogenous variable. We apply the IV framework to our study using a very simple measurement error model. The Online Appendix Section 6 shows the model together with the assumptions required to estimate the characteristics of compliers, never reporters and always reporters[22]. Never-reporters are respondents who report no IPV even when offered the ballot. Always reporters report IPV with the "Audible Response" method. Compliers are those who report IPV only when offered the ballot. The Appendix explains that, under the assumptions considered, the covariate means for compliers can be estimated residually from the whole sample.

Figure 1 plots covariate averages for always-reporters, never-reporters and compliers, alongside the overall mean for the entire sample. We only show the means graphically for the two covariates in which we observe significant differences across compliers and noncompliers – "no primary" and "illiterate" – and report averages for other covariates in Online Appendix Table A6.[23] As shown in the figure, compliers are much more likely to be illiterate and to not have completed primary school, suggesting that low-educated women were most likely to be marginal reporters of IPV in response to the increase in privacy offered with the ballot method. While there is some overlap in the confidence intervals, the p-values shown in Online Appendix Table A6 are statistically significant.[24] These patterns imply that, not only are average IPV rates undermeasured using existing survey methods, but they are disproportionately so for lower-SES households, so fail to adequately capture the relationship between interpersonal violence and poverty.[25]

[20] The three questions were: "Are you Peruvian?", "Is your mother Peruvian?", and "Have you attended college?"

[21] For the first two questions, the correct answer was the same for every woman in our sample by construction. The third question was asked twice, once with the traditional method and once using the ballot method.

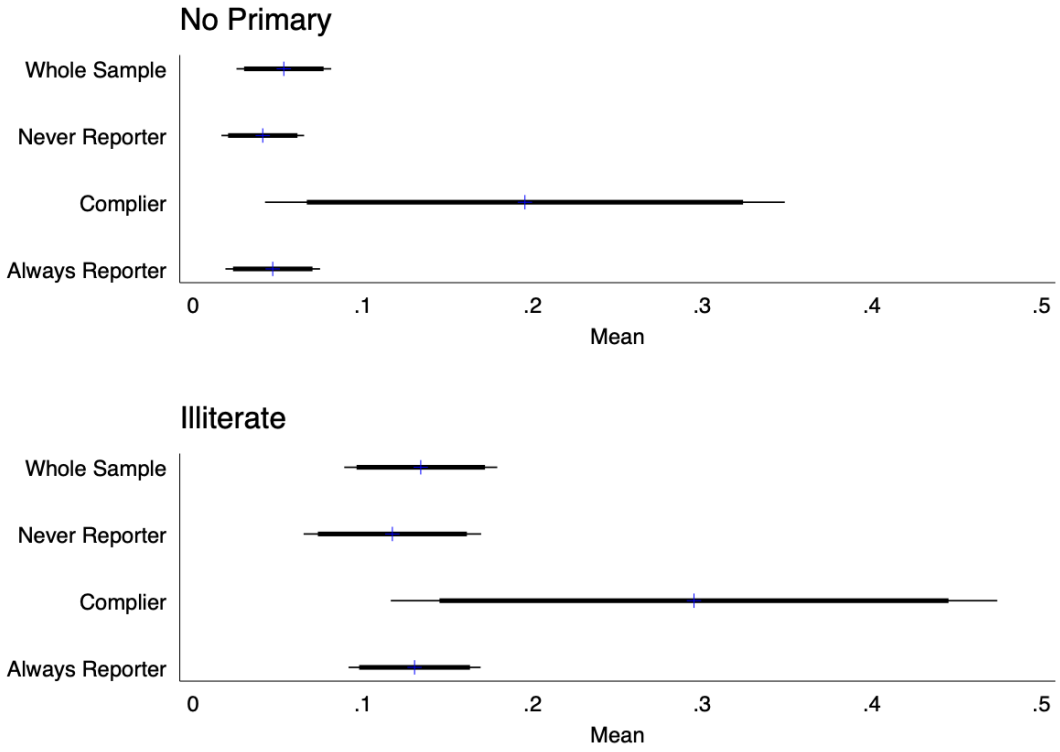
[22] The derivation follows closely the approach proposed by Marbach and Hangartner [2020].

[23] "No primary" is a dummy for not completing primary school, while "illiterate" indicates the respondent cannot read or write.

[24] Very low educated women may prefer the ballot as added privacy is particularly valuable if they are at higher risk of retaliation or escalation.

[25] Agüero and Frisancho [2022] conduct a study on a sample of peri-urban women entrepreneurs in Lima and find no average difference between self-reports elicited via DHS and list experiment methods, nor across sixteen subgroups, with the exception of women with higher education, who report higher IPV under the list experiment.

Figure 1. Marginal Respondents Have Extremely Low Education Levels



Notes: These figures plot covariate means for the whole sample, never-reporters, compliers and always-reporters. No primary indicates the respondent did not finish primary school. Illiterate are respondents who can neither read nor write. 90% and 95% confidence intervals shown from 1000 clustered bootstrap replications.

4 How reducing under-reporting changes inference

We next consider the implications of reducing measurement error via the ballot method for empirical inference on the drivers of IPV. By comparing "Audible Response" and "Private Response" IPV data within the context of two distinct natural experiments in Peru, we find that incidents revealed under enhanced privacy fundamentally change the analytical conclusions regarding IPV determinants.

First, we use a natural experiment embedded in our data collection timeline to investigate how IPV responds to temporal variation in household financial stress. Whether household resource injections increase or decrease women's exposure to IPV is theoretically ambiguous, and there is conflicting evidence in the empirical literature.[26] In order to further explore the relationship between IPV and household resource levels, we make use of variation in the time since last payment received through Peru's large-scale cash transfer program for the poor, which reaches roughly one third of households in our sample.[27] Prior work on household consumption smoothing has demonstrated that financial stress varies over time in liquidity-

[26] For instance, Hsu [2017]; Carr and Packham [2021] find that heightened resources increase IPV, while Hidrobo et al. [2016]; Buller et al. [2018]; Heath et al. [2020] find the opposite pattern.

[27] *Juntos* was introduced by the government in 2005 and it currently makes transfers of approximately $60 every 60 days to households living below a fixed poverty threshold. In total, 35.68% of households in our sample claim to receive Juntos payments.

constrained populations with distance between welfare payment periods.[28] This predictable variation allows us to test whether household resource levels exacerbate or reduce IPV prevalence in our sample and compare the patterns that emerge when using IPV measured under privacy conditions versus IPV measured using standard survey approaches.

Figure 2a plots IPV trends in relation to time since transfer payment, comparing IPV measured by ballot versus the "Audible Response" method.[29] The exercise reveals important differences in IPV trends over the payment cycle across the two different survey approaches. In particular, analysis employing IPV data from "Audible Response" questioning indicates no significant relationship between household resource levels and the incidence of sexual IPV: reported levels of violence are roughly constant across the payment cycle, hovering at around 10% throughout. In contrast, the same analysis using data collected via the ballot method reveals strong cyclical variation in sexual violence that corresponds to the timing of transfer receipt, consistent with other empirical work in this vein. Specifically, the pattern suggests that episodes of violence increase at the time of transfer to around 20% - roughly doubling - and revert back to initial levels by the end of the payment cycle.[30]

The stark difference is also evident in the regression estimates shown in Online Appendix Table A7: the analysis of audible response data on IPV reveals no statistically significant relationship between time since payment and violence, whereas the analysis based on ballot measurement indicates that household financial stress explains a significant portion of variation in sexual IPV prevalence in our sample.[31] More generally, the fact that the rates reported across methods are roughly the same at Day 0 suggests that *transitory* episodes of violence are more frequently underreported than steady state levels of violence, which has important implications for the study of temporary or cyclical changes in circumstance on IPV. An important caveat is that the patterns may also reflect an interaction between household resource levels and marginal responses to privacy conditions (e.g. financial stress may lead respondents to more readily report IPV when offered privacy), which we cannot entirely rule out, but such a relationship is difficult to reconcile without corresponding changes in the severity of IPV.

Note that the potential to pick up cyclicality on a stock variable such as lifetime IPV only exists if there is imperfect recall characterized by recency bias in reported IPV, a phenomenon that has been documented in other samples such as Abramsky et al. [2022].[32] In future waves of data collection, we find evidence of similar levels of recency bias in our sample based on longitudinal discrepancies in reported lifetime IPV that correlate with recency and severity of violence, suggesting that lifetime IPV responses at baseline indeed have the potential to pick up cyclicality in our sample.

The importance of survey technique also emerges in an analysis of the impact of a

[28] Specifically, past research has shown that the poor often have difficulty smoothing consumption between pay checks [Aguila et al., 2017; Stephens Jr, 2003], giving rise to predictable changes in household resource levels over the pay period. Fernández and Saldarriaga [2014] show evidence of imperfect consumption smoothing studying the same cash transfer program in Peru.

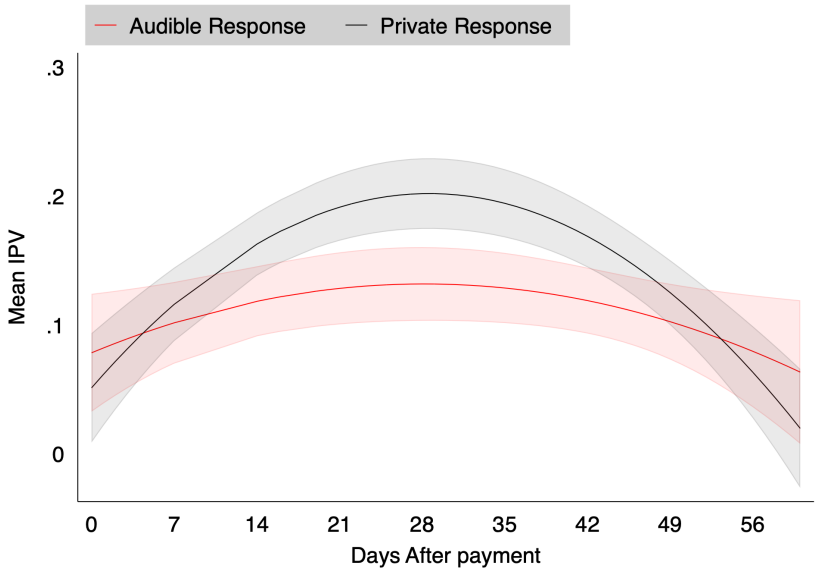
[29] Online Appendix Figure A2 shows that a similar pattern emerges using a non-parametric approach.

[30] Note that reported incidence does not jump discontinuously at 0 days after payment only because the question on violence pertains to a reference period of more than one day. Hence, the break in slope from decreasing to increasing that occurs at 0 days indicates an onset of heightened violence. As the reference period for violence reporting encompasses more post-payment days, violence continues to increase until it peaks at 4 weeks post payment, after which it begins to decline back to base levels.

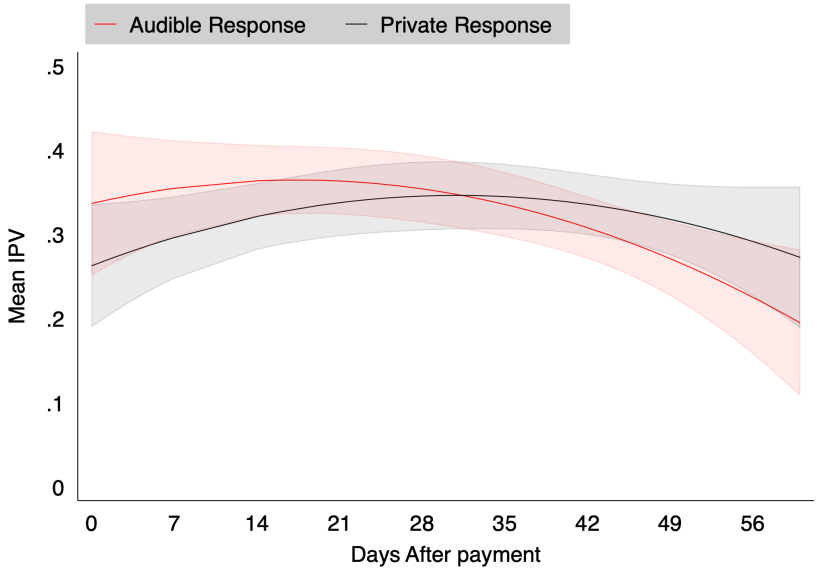
[31] As shown in columns 3-4 of Online Appendix Table A7 and also Figure 2b, there is no difference across reporting methods in the reported cyclicality of physical IPV.

[32] Recency bias is a cognitive error in retrospective reporting in which respondents over-report recent violent events and under-report, forget, or telescope earlier events. This leads to inflated estimates of current, short-term prevalence and inaccurate, lower estimates of historical or lifetime abuse.

Figure 2. Quadratic Regressions for Days Since *Juntos* Payment on IPV



(a) Sexual IPV



(b) Physical IPV

Note: This figure displays the results of sexual and physical IPV on days after last *Juntos* payment, evaluated separately for the "Private Response" and "Audible Response" survey methods. Standard errors are clustered at the village level. The sample is restricted to villages with a *Juntos* payment.

policy-induced change in schooling attainment on IPV. We replicate in our sample a natural experiment first used by Weitzman [2018], who exploited a 1993 educational reform in Peru that made high school compulsory for children aged 11 or younger, which discontinuously

increased years of education obtained by "exposed" relative to older (those aged 12+ at the time of the reform) cohorts. Using the age discontinuity as an instrument for schooling alongside IPV data from the Peruvian DHS, Weitzman [2018] estimates that one additional year of school is associated with an 8 percent reduction in physical IPV and a 10 percent reduction in sexual IPV.

Replicating the analysis with our sample allows us to compare estimates obtained using "Private Response" versus "Audible Response" measurement strategies. We find that the estimated effect of years of education on IPV depends critically on how the outcome is measured. Table 2 shows the results of estimating the effect of years of education on physical or sexual IPV, after instrumenting for years of education with a dummy indicator of whether the cohort was exposed to the reform with a fuzzy Regression Discontinuity Design (RDD).[33] Just as in Weitzman [2018], we observe a negative association between years of education and IPV in our sample using the "Audible Response" approach to data collection (columns 1 to 3). However, when IPV is measured using the ballot method (column 4 to 6), we can no longer reject the hypothesis of a null effect of education on IPV.

If the exclusion restriction holds, this finding indicates that some women are pushed into underreporting, when using standard survey techniques, *as a result of the education reform*.[34] This could occur, for example, if education (or SES more generally) increases the stigma of IPV, which may be especially true for marginal school-goers who are more likely to come from poverty.[35] The more general lesson highlighted by this example is that underreporting is particularly problematic when it has the potential to induce a correlation between measurement error in the outcome variable and the endogenous variable (education, in this case), yielding a biased and inconsistent estimate of the causal effect of education on true IPV. This concern has the potential to arise in many studies of determinants of stigmatized behaviors, on account of the fact that forces that influence the behavior are likely to also influence the stigma associated with the behavior, and hence rates of underreporting.

[33] We estimate a fuzzy RDD using a local linear polynomial within an MSE-optimal bandwidth. More details on the estimation can be found in Table 2. Online Appendix Table A8 shows a strong first stage, while Online Appendix Table A9 shows that we cannot reject the null hypothesis of density continuity around the cutoff based on the McCrary test, indicating that the design is valid.

[34] If, on the other hand, the exclusion restriction does not hold, the result instead indicates that some determinant of IPV underreporting other than years of schooling changes discontinuously at the cutoff.

[35] For instance, social desirability bias may increase with upward mobility [Agüero et al., 2026; Agüero and Frisancho, 2022; Cullen, 2023; Bursztyn et al., forthcoming].

Table 2. Effect of years of education on IPV (2SLS)

	Audible (DHS)			Private (Ballot)		
	(1) Physical	(2) Sexual	(3) Phys/Sex	(4) Physical	(5) Sexual	(6) Phys/Sex
Years of Education	-0.064**	-0.047**	-0.071**	0.022	-0.022	0.019
	(0.028)	(0.019)	(0.028)	(0.023)	(0.016)	(0.023)
Control Mean	0.467	0.146	0.487	0.509	0.176	0.526
Control SD	0.499	0.354	0.500	0.500	0.381	0.500
Observations	690	690	690	736	736	736
Clusters	229	229	229	234	234	234
F-statistic	29.19	29.19	29.19	32.68	32.68	32.68
Bandwidth (left)	4.86	4.86	4.86	5.75	5.75	5.75
Bandwidth (right)	4.86	4.86	4.86	5.75	5.75	5.75

Note: Fuzzy RD (2SLS) estimates using the MSE-optimal bandwidth from the first-stage regression. Controls include language. Standard errors are clustered at the village level.

The above comparisons illustrate that non-random measurement error in standard survey data on IPV can easily lead to biased causal estimates of IPV determinants, even when employing arguably exogenous variation in those factors. Moreover, these patterns also tell us how reporting bias varies with respondent conditions. For instance, the first estimates indicate that transitory incidents of IPV are more likely to be under-reported, in particular those driven by heightened resource levels. When violence is at its lowest in the cycle – right before payment receipt – incidence of violence is similar under the two reporting conditions. However, respondents interviewed without sufficient privacy disproportionately fail to report incidents of violence that arise in response to the cyclical increase in cash on hand. This is consistent with female respondents either feeling greater shame or stigma over violence driven by instrumental or manipulative incentives, or being more concerned about possible repercussions of reporting such events. As a result of this bias, estimates using IPV data collected with standard approaches are unable to identify household resource levels as a significant factor underlying patterns of household violence.

In the case of the natural experiment on education, we learn that underreporting potentially generates a spurious negative causal relationship between education and IPV because those who are pushed into additional schooling as a result of the policy reform become more likely to underreport IPV using standard survey instruments. This pattern, together with our analysis of "ballot compliers" in Section 3.2, showcases the complexity of predicting changes in IPV driven by factors such as education that are also likely to produce changes in reporting behavior. On the one hand, less educated individuals are more likely to underreport IPV, which hides some of the SES gradient in IPV incidence. At the same time, policies that push individuals into more schooling have the potential to further increase underreporting, which complicates causal inference on IPV responses to exogenous changes in SES.

5 Conclusions

This paper conducts a survey experiment to quantify the extent of measurement error in self-reported data on IPV. Despite high rates of self-reported prevalence, we find that IPV data

collected using gold-standard enumerator questioning are prone to a non-trivial degree of underreporting driven by privacy concerns. We show that simply allowing women to mark their answers privately through a ballot method increases reporting of violent acts, both physical and sexual, and that reporting improvements are largest for low-SES respondents. Meanwhile, ensuring a higher degree of privacy along with complete respondent anonymity by way of list experiment elicitation does not lead to higher reports of IPV compared to the ballot. Moreover, list experiments fail to generate individual-level data on IPV prevalence and produce noisier estimates.

Additionally, survey method matters beyond the accurate measurement of prevalence rates. Even when causal inference utilizes strictly exogenous variation in IPV drivers, privacy-motivated underreporting can lead researchers to both under- or over-estimate their influence. For women randomly selected to answer violence questions with more privacy, violence incidence is estimated to increase when cash payouts are received, consistent with the broad literature on liquidity constraints and stress, while such a pattern is not found among women that are asked about IPV with the traditional method. Meanwhile, evaluating the impact of an education reform that resulted in an increase in years of education shows a reduction in IPV using standard survey data that vanishes when employing IPV data collected under greater privacy. In general, these differences point to a critical need to address non-random measurement error in sensitive self-reported data in order to obtain unbiased estimates of IPV determinants.

References

Abramsky, Tanya, Sheila Harvey, Neema Mosha, Grace Mtolela, Andrew Gibbs, Gerry Mshana, Shelley Lees, Saidi Kapiga, and Heidi Stöckl, "Longitudinal inconsistencies in women's self-reports of lifetime experience of physical and sexual IPV: evidence from the MAISHA trial and follow-on study in North-western Tanzania," *BMC Women's Health*, 2022, *22* (1), 120.

Agüero, Jorge M and Veronica Frisancho, "Measuring violence against women with experimental methods," *Economic Development and Cultural Change*, 2022, *70* (4), 1565–1590.

Agüero, Jorge M., Viviana M.E. Perego, Julian Pedrazzi, and Javier Romero, "Privacy and Measurement Error in Phone Surveys: The Case of Women's Agency," 2026. Working paper.

Aguila, Emma, Arie Kapteyn, and Francisco Perez-Arce, "Consumption smoothing and frequency of benefit payments of cash transfer programs," *American Economic Review*, 2017, *107* (5), 430–435.

Ahlquist, John S, "List experiment design, non-strategic respondent error, and item count technique estimators," *Political Analysis*, 2018, *26* (1), 34–53.

Aizer, Anna, "The gender wage gap and domestic violence," *American Economic Review*, 2010, *100* (4), 1847–59.

_ , "Poverty, violence, and health the impact of domestic violence during pregnancy on newborn health," *Journal of Human Resources*, 2011, *46* (3), 518–538.

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin, "Identification of causal effects using instrumental variables," *Journal of the American statistical Association*, 1996, *91* (434), 444–455.

Barr, Anna Louise, Louise Knight, Ivan França-Junior, Elizabeth Allen, Dipak Naker, and Karen M Devries, "Methods to increase reporting of childhood sexual abuse in surveys: the sensitivity and specificity of face-to-face interviews versus a sealed envelope method in Ugandan primary school children," *BMC international health and human rights*, 2017, *17* (1), 4.

Benson, Lawrence E, "Studies in secret-ballot technique," *Public Opinion Quarterly*, 1941, *5* (1), 79–82.

Bishop, George F and Bonnie S Fisher, ""Secret ballots" and self-reports in an exit-poll experiment," *Public Opinion Quarterly*, 1995, *59* (4), 568–588.

Blair, Graeme, Alexander Coppock, and Margaret Moor, "When to worry about sensitivity bias: A social reference theory and evidence from 30 years of list experiments," *American Political Science Review*, 2020, *114* (4), 1297–1315.

_ **and Kosuke Imai**, "Statistical analysis of list experiments," *Political Analysis*, 2012, *20* (1), 47–77.

_ , **Winston Chou, and Kosuke Imai**, "List experiments with measurement error," *Political Analysis*, 2019, *27* (4), 455–480.

Buller, Ana Maria, Amber Peterman, Meghna Ranganathan, Alexandra Bleile, Melissa Hidrobo, and Lori Heise, "A mixed-method review of cash transfers and intimate partner violence in low-and middle-income countries," *The World Bank Research Observer*, 2018, *33* (2), 218–258.

Bursztyn, Leonardo, Ingar K. Haaland, Nicolas Röver, and Christopher Roth, "The Social Desirability Atlas," *Journal of Political Economy: Microeconomics*, forthcoming.

Cagney, Jack, Emmanuela Gakidou, Joht Singh Chandan, Felicia Marie Knaul, Nicholas Metheny, and Claudia Garcia-Moreno, "Women's retrospective reports of childhood sexual abuse in cross-sectional household surveys: a multi-country secondary analysis of two data collection methods," *The Lancet Global Health*, 2025, *13* (12), e2071–e2081.

Carr, Jillian B and Analisa Packham, "SNAP schedules and domestic violence," *Journal of Policy Analysis and Management*, 2021, *40* (2), 412–452.

Chowdhury, ME, N Alam, I Anwar, A Ahmed, T Saidel, PS Mallick, R Kelly, and PK Streatfield, "Assessment of non-marital sexual behaviours of men in Bangladesh: a methodological experiment using a modified confidential ballot-box method," *International journal of STD & AIDS*, 2012, *23* (3), e13–e17.

Chuang, Erica, Pascaline Dupas, Elise Huillery, and Juliette Seban, "Sex, lies, and measurement: Consistency tests for indirect response survey methods," *Journal of Development Economics*, 2021, *148*, 102582.

Coffman, Katherine B., Lucas C. Coffman, and Keith M. Marzilli Ericson, "The Size of the LGBT Population and the Magnitude of Antigay Sentiment Are Substantially Underestimated," *Management Science*, 2017, *63* (10), 3168–3186.

Crépon, Bruno, Ahmed Elsayed, and Jules Gazeaud, "Measuring Sensitive Outcomes through Ballot-Bag Surveys," *American Economic Review: Insights*, forthcoming.

Cullen, Claire, "Method matters: The underreporting of intimate partner violence," *The World Bank Economic Review*, 2023, *37* (1), 49–73.

De Cao, Elisabetta and Clemens Lutz, "Sensitive survey questions: Measuring attitudes regarding female genital cutting through a list experiment," *Oxford Bulletin of Economics and Statistics*, 2018, *80* (5), 871–892.

DeKeseredy, Walter S and Martin D Schwartz, "Measuring the extent of woman abuse in intimate heterosexual relationships: A critique of the Conflict Tactics Scales," *US Department of Justice Violence Against Women Grants Office Electronic Resources*, 1998.

Druckman, James N, Mauro Gilli, Samara Klar, and Joshua Robison, "Measuring drug and alcohol use among college student-athletes," *Social Science Quarterly*, 2015, *96* (2), 369–380.

Fernández, Fernando and Victor Saldarriaga, "Do benefit recipients change their labor supply after receiving the cash transfer? Evidence from the Peruvian Juntos program," *IZA Journal of labor & Development*, 2014, *3*, 1–30.

Gibson, Mhairi A, Eshetu Gurmu, Beatriz Cobo, María M Rueda, and Isabel M Scott, "Indirect questioning method reveals hidden support for female genital cutting in South Central Ethiopia," *PloS One*, 2018, *13* (5).

Gilligan, Daniel O, Melissa Hidrobo, Jessica Leight, and Heleene Tambet, "Using a list experiment to measure intimate partner violence: cautionary evidence from Ethiopia," *Applied Economics Letters*, 2025, *32* (11), 1594–1600.

Gregson, S, P Mushati, PJ White, M Mlilo, C Mundandi, and C Nyamukapa, "Informal confidential voting interview methods and temporal changes in reported sexual risk behaviour for HIV transmission in sub-Saharan Africa," *Sexually transmitted infections*, 2004, *80* (suppl 2), ii36–ii42.

Heath, Rachel, Melissa Hidrobo, and Shalini Roy, "Cash transfers, polygamy, and intimate partner violence: Experimental evidence from Mali," *Journal of Development Economics*, 2020, *143*, 102410.

Hidrobo, Melissa, Amber Peterman, and Lori Heise, "The effect of cash, vouchers, and food transfers on intimate partner violence: evidence from a randomized experiment in Northern Ecuador," *American Economic Journal: Applied Economics*, 2016, *8* (3), 284–303.

Hsu, Lin-Chi, "The timing of welfare payments and intimate partner violence," *Economic Inquiry*, 2017, *55* (2), 1017–1031.

Iqbal, Nasir, Amen Jalal, Mahreen Mahmud, and Kate Vyborny, "Collecting Accurate Data on Intimate Partner Violence: Learnings from Pakistan," Policy Research Working Paper 11077, World Bank, Washington, DC 2025.

Karlan, Dean S and Jonathan Zinman, "List randomization for sensitive behavior: An application for measuring use of loan proceeds," *Journal of Development Economics*, 2012, *98* (1), 71–75.

Kishor, Sunita, "Domestic Violence Measurement in the Demographic and Health Surveys: The History and the Challenges," *Division for the Advancement of Women*, 2015.

Lépine, Aurélia, Carole Treibich, and Ben d'Exelle, "Nothing but the truth: Consistency and efficiency of the list experiment method for the measurement of sensitive health behaviours," *Social Science & Medicine*, 2020, *266*, 113326.

Marbach, Moritz and Dominik Hangartner, "Profiling compliers and noncompliers for instrumental-variable analysis," *Political Analysis*, 2020, *28* (3), 435–444.

Meyer, Bruce D, Wallace KC Mok, and James X Sullivan, "Household surveys in crisis," *Journal of Economic Perspectives*, 2015, *29* (4), 199–226.

Olken, Benjamin A, "Monitoring corruption: evidence from a field experiment in Indonesia," *Journal of political Economy*, 2007, *115* (2), 200–249.

Palermo, Tia, Jennifer Bleck, and Amber Peterman, "Tip of the iceberg: reporting and gender-based violence in developing countries," *American Journal of Epidemiology*, 2014, *179* (5), 602–612.

Park, David Sungho, Shilpa Aggarwal, Dahyeon Jeong, Naresh Kumar, Jonathan Robinson, and Alan Spearot, "Private but misunderstood? Evidence on measuring intimate partner violence via self-interviewing in rural Liberia and Malawi," *The World Bank Economic Review*, 2024, p. lhae040.

Peterman, Amber, Malick Dione, Agnes Le Port, Justine Briaux, Fatma Lamesse, and Melissa Hidrobo, "Disclosure of violence against women and girls in Senegal," *The World Bank Economic Review*, 2024, p. lhae039.

Phillips, Anna Elizabeth, John Molitor, Marie Claude Boily, Catherine M Lowndes, Kaveri Gurav, James Blanchard, and Michel Alary, "Informal confidential voting interviewing in a sexual risk assessment of men who have sex with men (MSM) and transgenders (hijra) in Bangalore, India," *Sexually transmitted infections*, 2013, *89* (3), 245–250.

Singh, Susheela, Lisa Remez, and Alyssa Tartaglione, *Methodologies for estimating abortion incidence and abortion-related morbidity: a review.*, Guttmacher Institute, 2010.

Stephens Jr, Melvin, ""3rd of tha month": do social security recipients smooth consumption between checks?," *American Economic Review*, 2003, *93* (1), 406–422.

Straus, Murray A, Sherry L Hamby, Sue Boney-McCoy, and David B Sugarman, "The revised conflict tactics scales (CTS2) development and preliminary psychometric data," *Journal of Family Issues*, 1996, *17* (3), 283–316.

von Russdorf, Sophie, Laura Ahlborn, Alessandra Hidalgo-Arestegui, Gerald McQuade, and Marta Favara, "A sound methodology: Measuring experiences of violent conflict through audio self-interviews," *Economics Letters*, 2024, *242*, 111879.

Weitzman, Abigail, "Does increasing women's education reduce their risk of intimate partner violence? Evidence from an education policy reform," *Criminology*, 2018, *56* (3), 574–607.