


## Using a list experiment to measure intimate partner violence: cautionary evidence from Ethiopia


Daniel O. Gilligan, Melissa Hidrobo, Jessica Leight & Heleene Tabet

**To cite this article:** Daniel O. Gilligan, Melissa Hidrobo, Jessica Leight & Heleene Tabet (31 Jan 2024): Using a list experiment to measure intimate partner violence: cautionary evidence from Ethiopia, Applied Economics Letters, DOI: [10.1080/13504851.2024.2308579](https://doi.org/10.1080/13504851.2024.2308579)

**To link to this article:** <https://doi.org/10.1080/13504851.2024.2308579>

 [View supplementary material](#) 

---

 [Published online: 31 Jan 2024.](#)

---

 [Submit your article to this journal](#) 

---

 [View related articles](#) 

---

 [View Crossmark data](#) 

---



## Using a list experiment to measure intimate partner violence: cautionary evidence from Ethiopia

Daniel O. Gilligan, Melissa Hidrobo, Jessica Leight and Heleene Tabet

Poverty, Gender and Inclusion unit, International Food Policy Research Institute, Washington, DC, USA

### ABSTRACT

While indirect response methods are increasingly used in surveys to measure sensitive behaviours such as intimate partner violence, important questions persist around how respondents understand and react to these methods. This article presents evidence from a list experiment measuring multiple forms of intimate partner violence in rural Ethiopia. We find that the list experiment does not generate estimates of intimate partner violence that are higher than direct response questions; rather, prevalence estimates using the list experiment are lower vis-à-vis prevalence estimates using the direct reports, and are sometimes even negative. We provide suggestive evidence that this pattern may reflect ‘fleeing’ behaviour by respondents who do not wish to be associated with statements indicating their exposure to intimate partner violence.

### KEYWORDS

List experiment; intimate partner violence; Ethiopia; measurement

### JEL CLASSIFICATION

O15; O10

### 1. Introduction

Intimate partner violence (IPV) is a major global health challenge. However, measurement of sensitive behaviours such as IPV is challenging (Peterman et al. 2018), and researchers have increasingly explored the use of list experiments and other indirect methods.<sup>1</sup> A list experiment measures the prevalence of a sensitive behaviour by randomizing respondents to receive a ‘control’ list of non-sensitive ‘baseline’ statements, or a ‘treatment’ list including the baseline list and a statement capturing the sensitive behaviour; the difference in means between the two list groups reveals the prevalence rate for the sensitive behaviour. If social desirability bias reduces reporting of IPV, the list experiment will generate higher (and presumably more accurate) estimates of violence (Hinsley et al. 2019).

In this article, we report on a survey conducted in Ethiopia that used direct elicitation methods as well as multiple list experiments (LE) to measure the prevalence of different domains (emotional, physical, and sexual) of IPV. We find that estimates of IPV prevalence constructed using list experiments are lower and in some cases negative, and provide

exploratory evidence that this may reflect ‘fleeing behaviour’ on the part of respondents.

Our results add to a growing body of evidence suggesting that the differences between list experiments and direct response questions employed to measure IPV vary depending on the context. In Peru and Ethiopia, similar prevalence rates of IPV were found with list experiments and direct questions (Agüero and Frisnacho 2021; Kotsadam and Villanger 2022). In contrast, other recent papers found LEs generated higher estimates of reported IPV (Bulte and Lensink 2019; Castilla and Murphy 2023; Cullen 2023; Lépine, Treibich, and D’Exelle 2020; Traunmüller, Kijewski, and Freitag 2018).

The mixed performance of list experiments may reflect the salience of challenges that have been identified; in particular, the use of list experiments is valid only under certain assumptions that include no design, floor, or ceiling effects, discussed in further detail below (Blair and Imai 2012). List experiments also generate less precise estimates, require large samples, and are not effective for very rare behaviours (Hinsley et al. 2019). Moreover, participants may not trust or fully comprehend LEs (Chuang et al. 2021). In conjunction

with this literature, we provide some suggestive evidence around the respondent and list characteristics that may lead to poor LE performance.

## II. Methods

### Survey design

The survey experiment was conducted between July and October 2019 in Oromia and Amhara, Ethiopia as part of a larger randomized controlled trial. Fifty five percent of sampled households (totalling 4,062 households) were randomized to receive the list questions, and 45% were randomized to not receive them.<sup>2</sup> Those randomized to the list group were further assigned to one of four lists: emotional violence (List A); physical violence (List B); sexual violence (List C); and control (List D). Randomization was conducted at the household level and stratified at the district (woreda) level. All eligible women, regardless of whether they received the list questions, were administered the direct IPV questions employing the modified conflict tactic scale (see the Appendix for more details).

Respondents randomized to receive the list questions were exposed to three different control lists including four statements (see Appendix 1). Those randomized to List A, B, or C received a fifth question that was different for each of the three lists, in order to elicit disclosure about multiple forms of violence within each domain of IPV. The control lists consisted of questions probing whether the respondent engaged in non-controversial infrequent activities such as attending a wedding. However, two of the three control lists included one sensitive question: list one posed a question about forced marriage, and list two posed a question about female genital mutilation. List three included no sensitive items. The inclusion of some sensitive items in the control list serves to ‘camouflage’ the researcher’s intent in measuring IPV (Chuang et al. 2021). Respondents were read each list at least twice and asked to report

the number of events that experienced in the last 13 months.<sup>3</sup> The list materials were not provided in writing given the low literacy level in the population.

The list experiment was conducted on 2,122 households across 13 districts in Oromia and Amhara.<sup>4</sup> The sample is drawn from households participating in the government’s Productive Safety Net Program with a child 0–36 months old, sampled as part of a larger evaluation. More details about the sample are provided in the Appendix.

### Estimation

We estimate the prevalence of IPV from the list experiment by regressing the number of affirmative list items on a binary variable for assignment to List A, List B or List C, as well as district fixed effects; standard errors are clustered at the subdistrict (kebele) level.<sup>5</sup> The resulting coefficient ( $\beta_{LE}$ ) represents the prevalence of IPV from the list experiment and is reported in conjunction with the estimated prevalence using direct questions ( $\beta_{DR}$ ); the difference; and the p-value corresponding to the test  $\beta_{DR} = \beta_{LE}$ . We also report the p-value corresponding to a chi squared test of the joint hypothesis that  $\beta_{DR} = \beta_{LE}$  across all nine items, as well as for each domain of violence. In Table A1, we present specification tests around the list experiment, noting that there is little evidence of ceiling and floor effects.<sup>6</sup>

## III. Results

### Summary statistics

Table 1 reports summary statistics for the sample randomized to list groups A, B, C, and control, including the pairwise differences in means and associated p-values.

### List method versus direct method

Table 2 reveals that the rate of directly reported past-year experience of IPV is between 3% and

<sup>2</sup>The intention was to randomize with an assignment of 50% to each arm, but due to a coding error, the realized proportions were slightly different.

<sup>3</sup>The Ethiopian calendar includes 13 months within a year.

<sup>4</sup>These districts and subdistricts were chosen by the non-governmental organization (World Vision) that is a partner in the broader evaluation of a graduation model program.

<sup>5</sup>We also show in Table A11 the same results estimated without clustered standard errors, and find a consistent pattern.

<sup>6</sup>For list 1 as enumerated in Appendix 1, the p-value associated with this test is .364; for list 2, the p-value is .626; for list 3, the p-value is .518.

**Table 1.** Summary statistics (mean or percentage) and balance.

	List A	N	List B	N	List C	N	List D (control)	N	Dif: A vs. control	p-value: A vs. control	Dif: B vs. control	p-value: B vs. control	Dif: C vs. control	p-value: C vs. control
Age of primary female	30.955	359	31.06	366	31.424	330	30.826	340	0.129	0.811	0.234	0.662	0.598	0.286
Primary female has some education	0.258	357	0.269	361	0.264	326	0.286	339	-0.028	0.415	-0.017	0.575	-0.022	0.517
Age of primary male	39.155	349	39.209	358	39.183	323	38.827	329	0.328	0.668	0.383	0.611	0.356	0.601
Primary male has some education	0.416	346	0.398	357	0.425	322	0.473	328	-0.056	0.126	-0.075	0.040	-0.047	0.225
Household size	6.259	359	6.092	369	6.311	331	6.232	340	0.027	0.840	-0.14	0.309	0.079	0.576
Female-headed household	0.095	359	0.089	369	0.073	330	0.103	340	-0.008	0.721	-0.014	0.554	-0.03	0.206
Household head's main activity is crop production	0.896	356	0.907	365	0.921	329	0.891	340	0.005	0.829	0.016	0.450	0.030	0.137
Experienced emotional violence in the past 13 months	0.125	359	0.089	369	0.121	331	0.115	340	0.011	0.639	-0.025	0.266	0.006	0.803
Experienced physical violence in the past 13 months	0.109	359	0.051	369	0.076	331	0.068	340	0.041	0.029	-0.016	0.391	0.008	0.705
Experienced sexual violence in the past 13 months	0.033	359	0.024	369	0.057	331	0.038	340	-0.005	0.708	-0.014	0.293	0.019	0.203
Experienced physical or sexual violence in the past 13 months	0.125	359	0.070	369	0.106	331	0.082	340	0.043	0.028	-0.012	0.58	0.023	0.303
A husband is not justified in beating his wife in any of these situations (female response)	0.446	354	0.441	365	0.410	329	0.419	339	0.027	0.418	0.022	0.578	-0.009	0.830
A husband is not justified in beating his wife in any of these situations (male response)	0.682	321	0.633	338	0.624	298	0.655	296	0.027	0.498	-0.022	0.578	-0.031	0.436

This table reports summary statistics by randomization arm and the p-values corresponding to tests comparing each list experiment treatment arm to the control arm.

**Table 2.** Analysis of list experiment.

	Direct report	N	List report	N	Difference	p-value
Emotional violence						
Emotional violence: Said something to humiliate you in front of others	0.041 (0.005)	1399	0.018 (0.081)	699	0.022	0.782
Emotional violence: Threatened to hurt or harm you or someone you care about	0.02 (0.004)	1399	-0.037 (0.079)	698	0.057	0.466
Emotional violence: Insulted you or made you feel bad about yourself	0.103 (0.009)	1399	0.045 (0.083)	698	0.057	0.493
Physical violence						
Physical violence: Pushed you, shook you, or thrown something at you	0.038 (0.005)	1399	0.067 (0.082)	709	-0.029	0.725
Physical violence: Slapped you	0.054 (0.007)	1399	0.013 (0.080)	709	0.04	0.613
Physical violence: Twisted your arm or pulled your hair	0.026 (0.005)	1399	-0.031 (0.069)	708	0.057	0.493
Sexual violence						
Sexual violence: Physically forced you to have sexual intercourse	0.034 (0.005)	1399	0.008 (0.088)	670	0.026	0.765
Sexual violence: Physically forced you to perform any other sexual acts	0.022 (0.004)	1397	-0.026 (0.085)	669	0.049	0.563
Sexual violence: Forced you with threats/in another way to perform sexual acts	0.016 (0.004)	1398	-0.065 (0.077)	670	0.082	0.283
Joint (chi-squared) test across all nine items						0.363
Joint test: emotional violence items						0.427
Joint test: Physical violence items						0.524
Joint test: Sexual violence items						0.397

The direct report column reports the prevalence estimate for the specified act of violence in the last 13 months derived from the direct report questions; the list experiment column reports the prevalence estimate for the specified act derived from an analysis of the list experiment question, in which the respondent's answer is regressed on a binary variable for treatment list, conditional on district fixed effects and with standard errors clustered at the sub-district level. The final column reports the p-value for a test of equality across the direct report and list experiment estimates. The final row reports a joint test of equality across the direct report and list experiment questions, pooling across all nine items. Asterisks indicate significance at the ten, five, and one percent level.

8%. Using the list experiment, we observe prevalence rates that are close to zero or, when positive, similar to the direct report estimates. We cannot reject the hypothesis that  $\beta_{DR} = \beta_{LE}$  for any of the nine forms of IPV measured, again including forms of emotional, physical and sexual IPV,<sup>7</sup> though the failure to reject this hypothesis may substantially reflect imprecision in the list experiment estimates. Results are similar when estimated conditional on demographic controls (Table A2), estimated using the simple mean difference across the control and treatment lists (Table A3), estimated with a Poisson model (Table A4), and estimated restricting the data used to estimate the direct prevalence rate to the control arm only (Table A5).<sup>8</sup>

In addition, four out of nine prevalence rates estimated using the list experiment are negative, suggesting that on average, respondents presented with the treatment list including the additional sensitive, IPV-related item concurred with fewer items than respondents presented with the control list. One hypothesis is that respondents are 'fleeing' from the sensitive item in the treatment lists; they may experience cognitive overload when asked about stigmatized events, while simultaneously being asked to add up the total number of all events relevant to the list experiment (McEwan and Sapolsky, 1995). While we cannot directly substantiate this hypothesis, it is consistent with other evidence that respondent reactions to list experiments can

<sup>7</sup>Although these tests may be underpowered to detect differences across the direct response and list experiment under the assumption that reporting bias of sensitive IPV experiences attenuates the direct report estimates below the truth, the results suggest that the list experiment is no better than the direct response approach at estimating IPV prevalence.

<sup>8</sup>When we estimate the Poisson model, we calculate the p-values comparing across the direct and list experiment samples by testing the equality of the coefficient from the Poisson model to the simple prevalence rate from the direct question; we do not incorporate the standard error on this simple prevalence rate.

be unpredictable (Chuang et al. 2021; Kramon and Weghorst 2019). By contrast, low comprehension would not be expected to lead to negative prevalence rates, but rather to rates close to zero if respondents simply randomized their responses to both lists.

There is some evidence that the fleeing behaviour may be most pronounced for sexual IPV, where two of the three prevalence rates estimated are negative (compared to one of three prevalence rates in the domains of emotional and physical IPV). However, this may in part also reflect the fact that the prevalence rates for sexual IPV are lowest.

We further probe the hypothesis of ‘fleeing’ behaviour by investigating heterogeneity with respect to respondent, enumerator and list characteristics. In particular, we examine subsamples in which IPV might be more stigmatized and respondents less likely to admit experiencing violence; these include subsamples defined by age, education, self-efficacy,<sup>9</sup> whether the respondent states that IPV is justified, and enumerator sex. We also analyse different domains of violence, and different control lists. In particular we compare control lists with and without

a sensitive (non-IPV) question, as the former may camouflage the researchers’ interest in IPV (Chuang et al. 2021) and thus lead to less fleeing behaviour. For each subsample, we report the mean difference ( $\beta_{DR} - \beta_{LE}$ ) for each domain of violence in conjunction with the p-value corresponding to the test of the hypothesis that  $\beta_{DR} - \beta_{LE}$  equals zero.

Table 3 reveals no evidence of variation in fleeing behaviour across subgroups defined by educational status or their attitudes around whether IPV is justified,<sup>10</sup> and only weak evidence of heterogeneity by age and by enumerator sex. By contrast, we generally find that fleeing behaviour is concentrated among women characterized by high self-efficacy; these women may be more susceptible to feelings of shame when responding to a list experiment including IPV items. In Table 4, we report a parallel analysis by the control list employed, rather than the type of IPV elicited. We find some evidence that the pattern of negative prevalence rates is most pronounced for list three, where there is no sensitive control item. Again, we regard this as consistent with our hypothesis of fleeing behaviour, but acknowledge that the evidence is only suggestive.

**Table 3.** Analysis of list experiment by subsamples and type of violence.

	Obs.	Emotional violence		Physical violence		Sexual violence		Pooled violence	
		Average difference ( $\beta_{DR} - \beta_{LE}$ )	p-value: ( $\beta_{DR} - \beta_{LE}$ ) = 0	Average difference ( $\beta_{DR} - \beta_{LE}$ )	p-value: ( $\beta_{DR} - \beta_{LE}$ ) = 0	Average difference ( $\beta_{DR} - \beta_{LE}$ )	p-value: ( $\beta_{DR} - \beta_{LE}$ ) = 0	Average difference ( $\beta_{DR} - \beta_{LE}$ )	p-value: ( $\beta_{DR} - \beta_{LE}$ ) = 0
<b>Age</b>									
Under 30	1501	0.063	0.474	0.076	0.416	0.146	0.112	0.095	0.197
30 or older	2534	0.038	0.643	-0.018	0.824	-0.01	0.899	0.003	0.956
<b>Education</b>									
Any education	982	0.063	0.568	0.079	0.482	-0.088	0.416	0.018	0.835
No education	2843	0.03	0.672	-0.009	0.901	0.11	0.147	0.044	0.448
<b>Self-efficacy</b>									
Higher self-efficacy	2365	0.134	0.091*	0.147	0.068*	0.07	0.367	0.117	0.062*
Lower self-efficacy	1441	-0.128	0.243	-0.219	0.023	-0.028	0.796	-0.125	0.132
<b>Violence perceptions</b>									
Violence is never justified	1595	0.109	0.283	0.015	0.875	0.035	0.746	0.053	0.510
Violence is justified	2228	0.01	0.907	0.015	0.851	0.058	0.475	0.027	0.684
<b>Sex of Enumerator</b>									
Female enumerator	2154	-0.003	0.972	0.000	0.996	-0.044	0.638	-0.016	0.830
Male enumerator	1098	0.07	0.380	0.015	0.862	0.076	0.373	0.054	0.425

This table reports the average difference between the estimated prevalence of IPV using a direct report and the estimated prevalence of IPV using a list experiment ( $\beta_{DR} - \beta_{LE}$ ) for each of a series of subsamples (defined at left) and for each domain of violence, including a pooled measure of violence (defined at top). For each difference reported, the p-value corresponding to the test of the hypothesis that  $\beta_{DR} - \beta_{LE}$  is equal to zero is also reported. Asterisks indicate significance at the ten, five, and one percent level.

<sup>9</sup>We use the General Self-Efficacy Scale as a proxy for self-efficacy.

<sup>10</sup>Estimates by subsample for each IPV question are reported in Tables A6-A10.



**Table 4.** Analysis of list experiment by subsamples.

	Obs.	Control list 1		Control list 2		Control list 3	
		Average difference ( $\beta_{DR} - \beta_{LE}$ )	p-value: ( $\beta_{DR} - \beta_{LE}$ ) = 0	Average difference ( $\beta_{DR} - \beta_{LE}$ )	p-value: ( $\beta_{DR} - \beta_{LE}$ ) = 0	Average difference ( $\beta_{DR} - \beta_{LE}$ )	p-value: ( $\beta_{DR} - \beta_{LE}$ ) = 0
<b>Age</b>							
Under 30	1501	0.036	0.735	0.145	0.134	0.104	0.272
30 or older	2534	-0.01	0.899	-0.02	0.806	0.041	0.628
<b>Education</b>							
Any education	982	0.03	0.817	-0.03	0.808	0.054	0.643
No education	2843	-0.008	0.913	0.078	0.286	0.061	0.394
<b>Self-efficacy</b>							
Higher self-efficacy	2365	0.032	0.689	0.136	0.117	0.184	0.023**
Lower self-efficacy	1441	-0.115	0.287	-0.12	0.261	-0.14	0.128
<b>Violence perceptions</b>							
Violence is never justified	1595	-0.011	0.919	-0.004	0.969	0.174	0.076*
Violence is justified	2228	0.028	0.755	0.092	0.294	-0.038	0.631
<b>Enumerator gender</b>							
Female enumerator	2154	-0.147	0.144	0.096	0.304	0.003	0.974
Male enumerator	1098	0.106	0.219	-0.005	0.954	0.061	0.431

This table reports the average difference between the estimated prevalence of IPV using a direct report and the estimated prevalence of IPV using a list experiment ( $\beta_{DR} - \beta_{LE}$ ) for each of a series of subsamples (defined at left) and for multiple control lists employed in the list experiment (noted at top). For each difference reported, the p-value corresponding to the test of the hypothesis that  $\beta_{DR} - \beta_{LE}$  is equal to zero is also reported. Asterisks indicate significance at the ten, five, and one percent level.

## IV. Conclusion

In this article, we document that a list experiment conducted in Ethiopia to measure the prevalence of IPV was not effective in eliciting higher reports of violence vis-à-vis direct questions. We present suggestive evidence that this may reflect ‘fleeing’ behaviour on the part of respondents.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

We gratefully acknowledge funding for the impact evaluation by USAID under Cooperative Agreement No AID-FFP-A-16-00008. This work was undertaken as part of the CGIAR Research Program on Policies, Institutions, and Markets (PIM) led by the International Food Policy Research Institute (IFPRI). Additional funding support for this study was also provided by PIM.

## References

Agüero, J. M., and V. Frisancho. 2021. “Measuring Violence Against Women with Experimental Methods.” *Economic Development and Cultural Change* 70 (4): 1565–1590. <https://doi.org/10.1086/714008>.

- Blair, G., and K. Imai. 2012. “Statistical Analysis of List Experiments.” *Political Analysis* 20 (1): 47–77. <https://doi.org/10.1093/pan/mpr048>.
- Bulte, E., and R. Lensink. 2019. “Women’s Empowerment and Domestic Abuse: Experimental Evidence from Vietnam.” *European Economic Review* 115:172–191. <https://doi.org/10.1016/j.euroecorev.2019.03.003>.
- Carolina, C., and D. M. Murphy. 2023. “Bidirectional Intimate Partner Violence: Evidence from a List Experiment in Kenya.” *Health Economics* 32 (1): 175–193.
- Chuang, E., P. Dupas, E. Huillery, and J. Seban. 2021. “Sex, Lies, and Measurement: Consistency Tests for Indirect Response Survey Methods.” *Journal of Development Economics* 148:102582. <https://doi.org/10.1016/j.jdeveco.2020.102582>.
- Cullen, C. A. 2023. “Method Matters: the Underreporting of Intimate Partner Violence.” *The World Bank Economic Review* 37 (1): 49–73.
- Hinsley, A., A. Keane, F. A. V. St John, H. Ibbett, A. Nuno, and A. Ellison. 2019. “Asking Sensitive Questions Using the Unmatched Count Technique: Applications and Guidelines for Conservation.” *Methods in Ecology and Evolution* 10 (3): 308–319. <https://doi.org/10.1111/2041-210X.13137>.
- Kotsadam, A., and E. Villanger. 2022. “Jobs and Intimate Partner Violence - Evidence from a Field Experiment in Ethiopia.” *Journal of Human Resources*.
- Kramon, E., and K. Weghorst. 2019. “(Mis) Measuring Sensitive Attitudes with the List Experiment: Solutions to List Experiment Breakdown in Kenya.” *Public Opinion Quarterly* 83 (S1): 236–263. <https://doi.org/10.1093/poq/nfz009>.
- Lépine, A., C. Treibich, and B. D’Exelle. 2020. “Nothing but the Truth: Consistency and Efficiency of the List Experiment

- Method for the Measurement of Sensitive Health Behaviours.” *Social Science and Medicine* 266:113326. <https://doi.org/10.1016/j.socscimed.2020.113326>.
- McEwen, B. S., and R. M. Sapolsky. 1995. “Stress and Cognitive Function.” *Current Opinion in Neurobiology* 5 (2): 205–216.
- Peterman, A., T. M. Palermo, S. Handa, and D. Seidenfeld. 2018. “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. <https://doi.org/10.1002/hec.3588>.
- Traunmüller, R., S. Kijewski, and M. Freitag. 2018. “The Silent Victims of Wartime Sexual Violence: Evidence from a List Experiment in Sri Lanka.” *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2937943>.