

Addressing Measurement Error in Intimate Partner Violence Self-report Data Using Multiple Overimputation and Multidimensional Quantitative Bias Analysis

 Irina Bergenfeld,^a Robin A. Richardson,^b  Alexandria R. Hadd,^a  Cari Jo Clark,^a Regine Haardörfer,^c Charis Wiltshire,^b  Timothy L. Lash,^b and  Angela M. Bengtson^b

Background: Intimate partner violence (IPV) is an important global health issue for which measurement error limits public health action. Although most national IPV prevalence estimates come from general health surveys like the Demographic and Health Surveys (DHS), such data probably underestimate prevalence compared with violence-focused surveys.

Methods: Using violence-focused surveys conducted in the same country and year (± 1) as validation data, we explored two methods of bias adjustment to address measurement error in DHS prevalence estimates. In multidimensional bias analysis, we directly adjusted summary prevalence estimates, using a range of possible sensitivities (10%–100%) and specificities (95%–100%) to elucidate their reasonable bounds. In multiple overimputation, we reestimated all IPV observations, incorporating prior information on measurement error, and averaged prevalence estimates over 50 iterations.

Results: Multidimensional bias analysis revealed that an assumption of 95% specificity resulted in negative prevalence estimates in some cases, confirming that false positives are likely negligible. Reasonable sensitivities varied considerably across countries and IPV types, likely due to differences in the number of items used to

assess IPV. Multiple overimputation-adjusted estimates were similar to survey estimates, except when unadjusted DHS estimates were <5% and highly discrepant. Past-year estimates were less discrepant than lifetime estimates, suggesting that recall bias may be a factor in underreporting.

Conclusion: This study examines measurement error due to IPV underreporting in specific contexts where external information exists, highlighting the need for more accurate IPV assessment using multiple items per domain and for internal validation studies to be incorporated into large-scale surveys.

Keywords: Bias analysis; Intimate partner violence; Measurement error; Underreporting

(*Epidemiology* 2025;36: 741–750)

Intimate partner violence (IPV) is a pervasive global health issue with negative impacts on victims, families, and communities. The World Health Organization (WHO) estimates that 27% of women and girls globally have experienced physical and/or sexual IPV, yet this statistic likely underestimates the true scope of the problem.¹ IPV is associated with a range of interrelated physical and mental health issues for victims, including depression, substance abuse, injuries, and reproductive health problems.² Moreover, children who witness IPV in the home are at higher risk of developmental problems and more likely to experience parental abuse and neglect, with impacts that may persist throughout the life course.^{3,4}

Despite its importance from an epidemiologic perspective, global knowledge about IPV prevalence is limited by challenges in measurement,⁵ including measurement error, generally, and underreporting, specifically. IPV prevalence has been estimated primarily through administrative data and self-report via survey questionnaires. Surveys using standard, behaviorally based items are considered the optimal means of assessment because they are not limited to those who seek legal or medical help and are, therefore, less subject to underestimation.⁶ However, self-report data are subject to measurement error due to ambiguous wording, poor administration, errors in recall, or participant fatigue.⁷ These limitations may reduce accuracy and bias estimates in either direction. For

Submitted February 9, 2025; accepted June 30, 2025

From the ^aDepartment of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA; ^bDepartment of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA; ^cDepartment of Behavioral, Social, Health, and Education Sciences, Rollins School of Public Health, Emory University, Atlanta, GA.

I.B.: Conceptualization; data curation; formal analysis; visualization; software; funding acquisition; methodology; writing—original draft. C.J.C., R.H., and T.L.L.: Writing—review and editing. R.A.R.: Writing—review and editing; conceptualization. A.R.H. and C.W.: Writing—review and editing; software. A.M.B.: Writing—review and editing; supervision; conceptualization.

This study was funded by the Eunice Kennedy Shriver Institute of Child Health and Human Development (Grant No. 1F31HD116541-01).

Disclosure: The authors report no conflicts of interest.

Underlying data are publicly available from the Demographic and Health Survey program upon request (<https://www.dhsprogram.com/data/available-datasets.cfm>). Code is available on figshare at 10.6084/m9.figshare.21669359.

 Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Irina Bergenfeld, Department of Global Health, Rollins School of Public Health, Emory University, 1518 Clifton Rd, Atlanta, GA 30322. E-mail: iberger@emory.edu.

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 1044-3983/25/366-741750

DOI: 10.1097/EDE.0000000000001896

sensitive topics such as IPV, participants are also more likely to refuse to answer questions or to respond untruthfully due to the high costs associated with disclosure, including stigma, shame, financial dependence on the abusive partner, and fear of reprisal.^{6,8,9}

Although there is no gold standard for assessing IPV,¹⁰ researchers can compare estimates from different sources to gain some sense of the direction and magnitude of bias. National monitoring of IPV prevalence is primarily accomplished by population-based multicomponent surveys (e.g., Demographic and Health Surveys, or DHS), which are administered at regular intervals but typically result in lower prevalence estimates than surveys specifically designed to study violence.⁶ These differences may be due to the amount and/or quality of training of survey enumerators to handle sensitive topics,¹⁰ item sets used to measure IPV, or survey methodology.¹¹ Although experts in violence against women acknowledge that rigorously designed violence-focused surveys likely produce more accurate statistics than general health surveys such as the DHS, a lack of resources means that these are typically only conducted as one-off studies in a limited number of countries.^{6,12,13} However, studies comparing face-to-face questionnaires with more anonymous data collection methods suggest that substantial underreporting may be present even in violence-focused surveys.^{14–16} In list experiments, for example, control participants receive a set of neutral statements, while experimental participants receive the neutral statements plus an additional sensitive statement. Researchers calculate the prevalence of underreported behaviors as the difference in the average number of statements endorsed by control and experimental groups.^{15,16} While these studies offer some information about the underreporting of violence, key limitations include a lack of generalizability to the population level, a lack of transportability, and a lack of methodological comparability. Such studies have produced a range of disparate findings on the degree of IPV underreporting in face-to-face surveys, with two studies finding no evidence of underreporting^{11,17} and several finding substantial underreporting of at least one form of IPV.^{15,16,18–20}

Comparatively little is known about IPV overreporting versus underreporting. It is generally assumed that false positives are rare due to the negative social consequences of IPV disclosure,^{6,8,9} yet there is little empirical evidence to either support or challenge this assumption. Two studies have examined overreporting of physical IPV among university students in Australia and the US.^{21,22} These studies suggest that despite overreporting of acts considered accidental or not taken seriously by either partner, specificity of the physical IPV items used in the DHS is likely greater than 95%. However, limitations of these studies warrant caution, given that the population of young adults in high-income countries is quite different than populations studied in the DHS (women 15–49 in low- and middle-income countries), only physical IPV was considered,

and underreporting was not studied concurrently, precluding the calculation of exact sensitivities and specificities.

Obtaining precise and accurate information about the prevalence of IPV is important for monitoring by national and international interested parties and for evaluating efforts to address IPV. During the past decade, numerous studies have been conducted to test various approaches to prevent violence against women, including social norms change,²³ economic empowerment,²⁴ and cash transfers.²⁵ However, without accurate IPV assessment, the findings of these studies may not be valid.¹⁹

Due to the uncertainty surrounding IPV underreporting and the lack of a gold standard, we explored two methods for bias adjustment of prevalence estimates from multicomponent surveys, benchmarking bias-adjusted estimates against estimates from violence-focused surveys conducted in the same country and year. The first method was multiple overimputation, an extension of multiple imputation that probabilistically imputes all observations of a variable based on prior information and existing data.²⁶ The second was multidimensional bias analysis, in which we assumed several different values of sensitivity and specificity and directly adjusted prevalence estimates to gain a sense of their reasonable bounds after accounting for potential bias.²⁷ Both methods are useful when there is little information about the true extent of bias.

METHODS

This study was determined to be exempt by the Institutional Review Board at Emory University.

Data and Sample

The sample for this investigation consists of all DHS conducted in the same country as and within 1 year of a population-based violence-focused survey, of which there were four total (Table 1). DHS are population-based health surveys conducted in low- and middle-income countries through the US Agency for International Development. DHS uses standard items to measure physical, sexual, and emotional IPV (ever and in the past 12 months) in women aged 15–49. IPV prevalence is the proportion of individuals within the sample of ever-partnered women who responded that their current or former partner engaged in at least one of the behaviors listed (1 = any IPV reported, 0 = no IPV reported). In the standard DHS module, this list includes seven acts of physical violence, three acts of sexual violence, and three acts of emotional abuse; however, some surveys have administered modules with fewer items (Supplemental Digital Content, eTable 1; <https://links.lww.com/EDE/C260>). For example, the Albania 2018 DHS assessed IPV using a single item: “Did a husband/partner ever hit you with his fists, kick you or do anything to hurt you physically?” In general, the physical acts assessed in violence-focused surveys and DHS with multiple

TABLE 1. Countries With Violence-focused Surveys Conducted in the Same Year as DHS

DHS Country and Year	No. Items by IPV Type, DHS	Violence-focused Survey, Year Conducted	Population Sampled, Violence-focused Survey ^a	No. Items by IPV Type, Violence-focused Survey
Peru, 2000 ^b (n = 18,764)	Lifetime physical (1)	WHO Multicountry Study on Women’s Health and Domestic Violence, 2000 ²⁸	Women/girls ages 15–49 who had ever married, cohabitated with, or had a regular sexual partnership with a man (n = 1,837)	Physical (6)
Egypt, 2014 (n = 6,693)	Lifetime and past year physical (7), emotional (3), and sexual (3)	Economic Cost of Gender-based Violence Survey, 2015 ²⁹	Ever-married women ages 18–64 (n = 18,100)	Physical (6), emotional (4), sexual (3)
Cambodia, 2014 (n = 3,499)	Lifetime and past year physical (7), emotional (3), and sexual (3)	National Survey on Women’s Health and Life Experiences in Cambodia, 2015 ³⁰	Ever-partnered (incl. dating partnerships) women/girls ages 15–64 ^c (n = 3,430)	Physical (6), emotional (4), sexual (3)
Albania, 2018 (n = 11,954)	Lifetime and past year physical (1)	OSCE-led Survey on Violence Against Women, 2019 ³¹	Ever-partnered (incl. dating partnerships) women ages 18–74 (n = 1,660)	Physical (9)
		National Population Survey, 2018 ³²	Ever-partnered women ages 18–74 (n = 3,314)	Physical (11)

^aDHS generally defines “partnership” as a cohabitating or married relationship with a man. In Egypt, only ever-married women were sampled.
^bFor comparability with WHO study, only estimates from rural areas were calculated.
^cEstimates for women 15–49 were reported and are used in this study.

items were the same, with some differences in sexual and emotional acts assessed across these surveys.

DHS enumerators receive training in survey administration, including ethical and safety guidelines, although the length of this training varies across surveys. All five violence-focused surveys (two for Albania and one each for Egypt, Cambodia, and Peru) used as comparators for DHS in this study were administered by the WHO and/or the United Nations and followed best practices for research on violence against women and girls, including training of enumerators to handle the topics with sensitivity and maintain privacy.³³ Moreover, all five surveys used similar multistage sampling procedures and standard, validated item sets.³⁴ Finally, violence-focused survey estimates used in this study are adjusted using sampling weights to be nationally representative.

Analytic Approach

Multiple Overimputation

Multiple overimputation is an extension of multiple imputation that simultaneously addresses missingness and measurement error by probabilistically imputing all observations of a mismeasured variable based on both prior information and available covariates.²⁶ This method has been applied to epidemiologic research involving error-prone measures such as HIV viral load³⁵ and gestational age³⁶ and is useful when the magnitude of bias is uncertain because no internal validation data are available.

Based on prior literature, we set different priors to account for overreporting and underreporting,^{6,22} hypothesizing the level of underreporting as the difference between

DHS and violence-focused survey estimates. We set priors for observations with IPV = 0 or missing between 0 and 0.5 (Figure 1), as overimputed values <0.5 (i.e., less than 50% probability of experiencing IPV) were recoded to 0, while overimputed values ≥0.5 were recoded to 1.³⁷ We then set wider confidence intervals around the priors for each model by dividing the difference in DHS and violence-focused survey prevalence (false negatives) by the proportion of women who reported no IPV in the DHS, then subtracting double this result from 100% to create a two-tailed distribution. The percentage of the distribution over 0.5 is the assumed percentage of false negatives over total negatives; the proportion less than 0.5 is the negative predictive value. If there was no evidence of underreporting (i.e., the DHS estimate was higher than/equal to the violence-focused survey estimate), a 98% confidence interval allowed for 1% of the distribution to be greater than 0.5 to account for noise. Due to a lack of internal information on overreporting of IPV and literature suggesting that false positives are relatively rare,²² we set priors for observations with IPV = 1 between 0.5 and 0.99 with a 98% confidence interval, resulting in 1% of the distribution being less than 0.5 (99% positive predictive value). Overimputed observations were bounded by 0 and 1, such that where the tails of the prior distributions were less than 0 or greater than 1, these were changed to 0 and 1, respectively.

In addition to physical, sexual, and emotional IPV, we included variables theoretically associated with IPV, with missingness, and with underreporting in our imputation models (e.g., age, education, literacy, household wealth, fear of spouse) (eTable 2; <https://links.lww.com/EDE/C260>). We used the R package *Amelia II* to produce 50 overimputed data

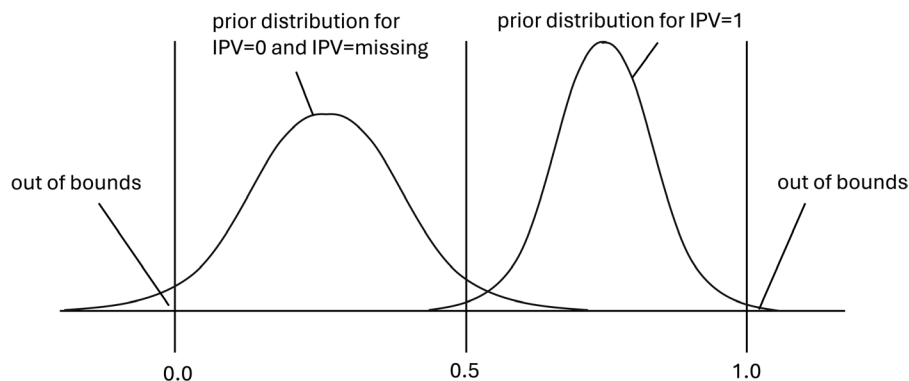


FIGURE 1. Example prior distributions used in the multiple overimputation procedure. The area under the curve for IPV = 0 and IPV = missing that is greater than 0.5 represents the hypothesized proportion of false negatives over total negatives. The area under the curve for IPV = 1 that is less than 0.5 represents the hypothesized proportion of false positives over total positives. Confidence intervals of the prior were thus calculated as $(100 - 2 \times [\text{false negatives}/\text{total negatives}])$. Full calculations are in Supplemental Digital Content, eTable 3; <https://links.lww.com/EDE/C260>.

sets³⁶ for each IPV type in each country. After dichotomizing all observations of overimputed IPV data, we calculated prevalence estimates for lifetime and past-year IPV, by type, for each of the 50 datasets. Using the *survey* package, we incorporated the survey weights, clusters, and strata provided by DHS into the prevalence calculations to produce bias-adjusted, nationally representative estimates (except for Peru, which was representative of rural areas only) and calculated the mean estimate across the 50 imputations.³⁸

Multidimensional Quantitative Bias Analysis

We performed a multidimensional bias analysis based on a range of possible sensitivities (Se) and specificities (Sp) for dichotomous IPV exposure. We evaluated specificities of 100%, 99%, and 95%, given the assumption that overreporting of IPV (false positive reporting) in these surveys is likely to be very low due to the stigma surrounding IPV.⁶ We evaluated sensitivities as high as 100% and as low as 10% in some cases where DHS and violence-focused survey estimates were highly discrepant, in intervals of 10%. For each country and type of IPV, we calculated a matrix of adjusted prevalence estimates for each combination of sensitivity and specificity using the formula:

$$\text{Adjusted prevalence} = (\text{Observed prevalence} + \text{Sp} - 1) / (\text{Se} + \text{Sp} - 1)$$

RESULTS

Contrary to expectation, not all estimates of IPV prevalence were less in DHS versus violence-focused surveys (Table 2). Although estimates for Albania and Peru – the two DHS that used single-item assessments for physical IPV – were substantially lower than violence-focused survey estimates, only lifetime sexual and emotional IPV were underestimated by >1% in the Cambodia DHS. For Egypt, all

TABLE 2. Demographic and Health Survey (DHS), Multiple Overimputation-Adjusted (MO), and Violence-focused Survey (VFS) Intimate Partner Violence Prevalence Estimates by Type

	Physical (%)		Emotional (%)		Sexual (%)	
	Lifetime	Past Year	Lifetime	Past Year	Lifetime	Past Year
Albania						
DHS	3.34	1.98	NA	NA	NA	NA
MO	11.80	3.32	NA	NA	NA	NA
VFS	18.0–19.0 ^a	7.2	NA	NA	NA	NA
Cambodia						
DHS	16.19	9.33	24.76	17.31	5.98	3.94
MO	15.52	9.78	28.89	13.82	7.67	4.38
VFS	15.5	5.7	32.0	14.7	9.8	4.7
Egypt						
DHS	25.24	13.53	18.80	13.07	4.13	2.72
MO	33.02	13.91	47.78	21.32	8.61	4.29
VFS	31.8	11.8	42.5	22.3	12.3	6.5
Peru						
DHS	40.71	NA	NA	NA	NA	NA
MO	64.54	NA	NA	NA	NA	NA
VFS	61.0	NA	NA	NA	NA	NA

^aRange of estimates from the two violence-focused surveys from Albania.

violence-focused survey estimates were substantially higher, except past-year physical IPV.

Multiple Overimputation

In general, multiple overimputation-adjusted DHS estimates were within one percentage point of violence-focused survey estimates, except where the two surveys were highly discrepant and the DHS estimate was <5%. For Albania, MO estimates using the prior distributions based on violence-focused survey estimates increased relative to DHS data but remained several percentage points below violence-focused survey prevalence estimates. For Peru, prior distributions

based on violence-focused survey data were too wide for the imputation model to converge and were reduced accordingly. Even with narrower prior distributions, multiple overimputation estimates overshot the violence-focused survey values by about four percentage points. For Cambodia, priors assigned based on violence-focused survey data produced prevalence estimates within one percentage point of violence-focused survey estimates for past-year emotional and sexual IPV but remained a few percentage points below lifetime violence-focused survey estimates for emotional and sexual IPV. For Egypt, multiple overimputation with priors based on violence-focused survey estimates overshot or came within a percentage point of all violence-focused survey physical and sexual IPV prevalence estimates. Lifetime and past-year sexual IPV, however, remained below the level of violence-focused survey prevalence after adjustment via multiple overimputation.

Multidimensional Bias Analysis

The bounds of reasonable sensitivity and specificity differed markedly by country, particularly for DHS surveys with single-item IPV assessment. In surveys using more items, probable sensitivities were markedly higher, especially for physical IPV (Figure 2).

Black Horizontal Lines Are Violence-focused Survey Prevalence Estimates

For Albania lifetime and past-year physical IPV, evaluating a specificity as low as 95% resulted in negative prevalence estimates (Table 3). Even assuming specificity to be 100% resulted in sensitivities of 20% or lower to achieve prevalence estimates for lifetime and past-year physical IPV similar to those from violence-focused surveys. For Peru, sensitivity for lifetime physical IPV in rural areas was between 60% and 70%, depending on the value

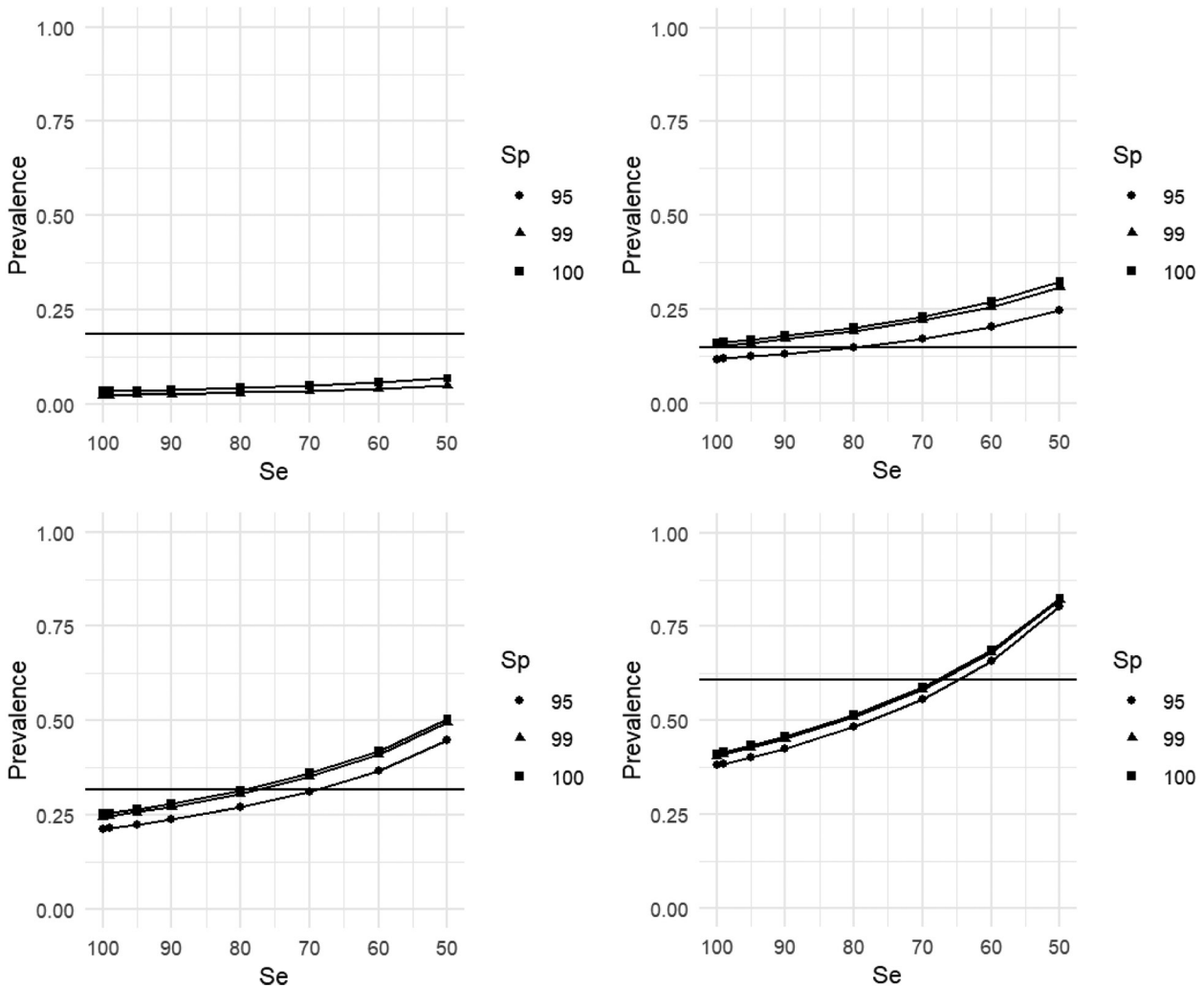


Figure 2. Adjusted prevalence estimates for lifetime physical intimate partner violence versus sensitivity (Se) at 95%, 99%, and 100% specificity (Sp) (clockwise from top left: Albania, Cambodia, Peru, and Egypt).

TABLE 3. Proportion of Individuals Exposed to Intimate Partner Violence (IPV) Under Different Sensitivities and Specificities, Albania 2018 DHS (n = 11,954)

Sp	Se									
	100%	90%	80%	70%	60%	50%	40%	30%	20%	10%
Lifetime physical IPV										
100%	0.034	0.038	0.043	0.049	0.057	0.069	0.086	0.114	0.172	0.343
99%	0.025	0.027	0.031	0.035	0.041	0.050	0.062	0.084	0.128	0.270
95%	-0.017	-0.018	-0.021	-0.024	-0.029	-0.035	-0.045	-0.063	-0.105	-0.314
Past-year physical IPV										
100%	0.020	0.022	0.025	0.028	0.033	0.040	0.050	0.066	0.099	0.198
99%	0.010	0.011	0.012	0.014	0.017	0.020	0.025	0.034	0.052	0.109
95%	-0.032	-0.036	-0.040	-0.046	-0.055	-0.067	-0.086	-0.121	-0.201	-0.604

TABLE 4. Proportion of Individuals Exposed to Intimate Partner Violence (IPV) Under Different Sensitivities and Specificities, Peru 2000 DHS (n = 18,764)

Sp	Se							
	100%	99%	95%	90%	80%	70%	60%	50%
Lifetime physical IPV								
100%	0.412	0.416	0.434	0.458	0.515	0.589	0.687	0.824
99%	0.406	0.410	0.428	0.452	0.509	0.583	0.682	0.821
95%	0.381	0.385	0.402	0.426	0.483	0.557	0.658	0.805

of specificity evaluated (Table 4). Physical IPV estimates in the Cambodia DHS were higher than those of the corresponding violence-focused survey (Table 5). Therefore, assuming a specificity $\geq 95\%$, the sensitivity of the DHS item set for physical IPV in Cambodia is between 80% and 100%. For sexual IPV, a specificity of 95% produced negative or extremely low prevalence estimates, while an assumed specificity of 99% or 100% suggested sensitivities between 60% and 70% for lifetime and between 70% and 80% for past-year IPV. For emotional IPV, past-year DHS estimates were higher, but lifetime estimates were lower. Assuming specificity $\geq 95\%$ suggested a sensitivity between 60% and 70% for lifetime emotional IPV, but over 90% for past-year emotional IPV. Egypt showed a similar pattern to Cambodia, but with lower sensitivities overall (Table 6). Lifetime physical IPV required a sensitivity of 70%–80% to match violence-focused survey estimates, whereas past year required sensitivity $>80\%$ for specificities between 95% and 99%. Sexual IPV estimates could not accommodate an assumption of 95% specificity without producing negative estimates. Under the assumption of $\geq 99\%$ specificity, sensitivity was between 30% and 40% for lifetime and between 40% and 50% for past-year sexual IPV. For emotional IPV, sensitivities of about 40% for lifetime and about 50% for past-year IPV were required to replicate estimates from violence-focused surveys.

TABLE 5. Proportion of Individuals Exposed to Intimate Partner Violence (IPV) Under Different Sensitivities and Specificities, Cambodia 2014 DHS (n = 3,499)

Sp	Se					
	100%	90%	80%	70%	60%	50%
Lifetime physical IPV						
100%	0.162	0.180	0.202	0.231	0.270	0.324
99%	0.153	0.171	0.192	0.220	0.257	0.310
95%	0.118	0.132	0.149	0.172	0.203	0.249
Past-year physical						
100%	0.093	0.104	0.117	0.133	0.156	0.187
99%	0.084	0.094	0.105	0.121	0.141	0.170
95%	0.046	0.051	0.058	0.067	0.079	0.096
Lifetime sexual						
100%	0.060	0.066	0.075	0.085	0.100	0.120
99%	0.050	0.056	0.063	0.072	0.084	0.102
95%	0.010	0.012	0.013	0.015	0.018	0.022
Past-year sexual						
100%	0.039	0.044	0.049	0.056	0.066	0.079
99%	0.030	0.033	0.037	0.043	0.050	0.060
95%	-0.011	-0.012	-0.014	-0.016	-0.019	-0.024
Lifetime emotional						
100%	0.248	0.275	0.310	0.354	0.413	0.495
99%	0.240	0.267	0.301	0.344	0.403	0.485
95%	0.208	0.232	0.263	0.304	0.359	0.439
Past-year emotional						
100%	0.173	0.192	0.216	0.247	0.289	0.346
99%	0.165	0.183	0.206	0.236	0.276	0.333
95%	0.130	0.145	0.164	0.189	0.224	0.274

DISCUSSION

This bias analysis of IPV prevalence data from general health surveys explored two analytic methods: multidimensional bias analysis and multiple overimputation. Multidimensional bias analysis revealed that an assumption of 95% specificity could not be accommodated in some cases (e.g., Albania physical IPV, Egypt sexual IPV) without resulting in negative prevalence estimates. In other cases, such as

TABLE 6. Proportion of Individuals Exposed to Intimate Partner Violence (IPV) Under Different Sensitivities and Specificities, Egypt 2014 DHS (n = 6693)

Sp	Se							
	100%	90%	80%	70%	60%	50%	40%	30%
Lifetime physical								
100%	0.252	0.280	0.316	0.361	0.421	0.505	0.631	0.841
99%	0.245	0.272	0.307	0.351	0.411	0.495	0.622	0.836
95%	0.213	0.238	0.270	0.311	0.368	0.450	0.578	0.810
Past-year physical IPV								
100%	0.135	0.150	0.169	0.193	0.226	0.271	0.338	0.451
99%	0.127	0.141	0.159	0.182	0.212	0.256	0.321	0.432
95%	0.090	0.100	0.114	0.131	0.155	0.190	0.244	0.341
Lifetime sexual IPV								
100%	0.041	0.046	0.052	0.059	0.069	0.083	0.103	0.138
99%	0.032	0.035	0.040	0.045	0.053	0.064	0.080	0.108
95%	−0.009	−0.010	−0.012	−0.013	−0.016	−0.019	−0.025	−0.035
Past-year sexual IPV								
100%	0.027	0.030	0.034	0.039	0.045	0.054	0.068	0.091
99%	0.017	0.019	0.022	0.025	0.029	0.035	0.044	0.059
95%	−0.024	−0.027	−0.030	−0.035	−0.041	−0.051	−0.065	−0.091
Lifetime emotional IPV								
100%	0.188	0.209	0.235	0.269	0.313	0.376	0.470	0.627
99%	0.180	0.200	0.225	0.258	0.302	0.363	0.456	0.614
95%	0.145	0.162	0.184	0.212	0.251	0.307	0.394	0.552
Past-year emotional IPV								
100%	0.131	0.145	0.163	0.187	0.218	0.261	0.327	0.436
99%	0.122	0.136	0.153	0.175	0.205	0.246	0.309	0.416
95%	0.085	0.095	0.108	0.124	0.147	0.179	0.231	0.323

lifetime sexual IPV in Cambodia, prevalence estimates under 95% specificity were lower than DHS estimates, even under assumptions of very low sensitivity. At least for these specific countries and IPV types, our findings suggest that overreporting of IPV is relatively rare and that DHS measures are highly specific. In general, past-year estimates had higher sensitivities than lifetime estimates for all types of IPV that were underreported, suggesting that recall bias may be a factor in underreporting. In particular, older women who have not experienced recent IPV may have either forgotten or normalized past experiences of violence.³⁹ Physical IPV tended to have higher sensitivities than other forms, likely due to the higher number of items and greater similarity of acts assessed in the DHS and violence-focused survey.

The multidimensional analysis also highlights key differences across countries: for example, the sensitivity of DHS items compared with a gold standard violence-focused survey was lower in Egypt than in Cambodia for the same types of IPV, despite using almost the same item sets and survey methodology. This may be due to differences in survey administration or to any number of cultural, political, or logistical factors. For example, Egypt DHS enumerators were recruited and trained over one week, while violence-focused survey enumerators in both countries were trained over two weeks

and likely had more specialized training in IPV disclosure. In contrast, Cambodia DHS enumerators received 26 days of training. It may also be the case that rapport-building was more necessary in Egypt due to recent civil unrest, lack of trust in government,⁴⁰ and/or a lack of coordination between government and nongovernmental organizations,⁴¹ which is necessary for the successful large-scale survey administration. However, we can only speculate about whether the larger discrepancies between DHS and violence-focused survey estimates in Egypt were the result of a more accurate violence-focused survey, a less accurate DHS, or a combination of both, and about the reasons behind these differences.

In most cases, multiple overimputation produced estimates similar to or greater than violence-focused survey estimates using prior distributions based on differences between DHS and violence-focused survey estimates. Bias adjustment using multiple overimputation resulted in estimates lower than those observed in violence-focused surveys in cases where the DHS estimates were both low (<5%) and highly discrepant from violence-focused surveys. When few individuals report IPV, the imputation model has less information about correlates of IPV and overimputed observations from individuals who did report IPV are less likely to change. For this reason, sexual IPV may be more difficult to bias-adjust with multiple

overimputation than other forms of IPV, as the reported prevalence of sexual IPV tends to be lower than other forms. This may be because sexual IPV is truly less prevalent than other forms of IPV, or because it is more likely to be underreported due to stigma.

Although it is commonly found that violence-focused surveys produce higher estimates than multicomponent surveys such as the DHS,⁶ our findings show that this is not universally true. In DHS that measured physical IPV using several items (Cambodia and Egypt), estimates are comparable to those from methodologically rigorous violence-focused surveys conducted in the same population and timeframe. DHS physical IPV estimates in our study were, however, substantially lower than violence-focused survey estimates when acts were bundled into a single question (as in Peru and Albania). In the two DHS that assessed emotional and sexual IPV, the estimates were more often lower than in comparable violence-focused surveys. This may be due to the two unique items present in violence-focused surveys. Gender-based violence experts have highlighted lack of content validity in the item sets used to measure emotional and sexual IPV; this study highlights the need for additional items to capture these domains of IPV. Social and cultural norms may also play a role in differential underreporting of different types of IPV. For example, list experiments conducted in Nigeria found underreporting of physical, but not sexual IPV,¹⁶ whereas the reverse was observed in Tanzania.¹⁸ Explanatory qualitative research conducted in the context of list experiments suggests that in some contexts, sexual IPV is considered less serious due to widespread beliefs about male sexual privilege. Likewise, physical IPV may be considered more normative in contexts where violence is considered an acceptable means of resolving conflict and maintaining household discipline.⁴²

Limitations

First, the true level of IPV underreporting is unknowable and likely greater than the difference in estimated prevalence between DHS and violence-focused surveys.⁶ Thus, the underreporting explored in this study may be considered a conservative estimate of false negatives in IPV research. Second, although we selected population-based violence-focused surveys conducted during the same period for comparison, some of these surveys defined their populations slightly differently than the DHS, which typically samples ever-partnered/married women ages 15–49. However, these sampling differences would generally be expected to produce lower violence-focused survey estimates than samples taken from the DHS. For example, women 50–74 and women in non-cohabitating relationships who were included in violence-focused surveys but absent from DHS samples would likely have lower risk of IPV compared with other groups, based on prior literature. Thus, actual differences in DHS and

violence-focused survey estimates can still be assumed to conservatively estimate underreporting. Third, we did not account for variation due to random chance. Fourth, we dichotomized continuous imputed IPV observations to calculate adjusted prevalences, which may result in some loss of information. Finally, we performed this study in a limited sample of diverse countries. Multidimensional bias analysis assumes that sensitivity and specificity are independent of true prevalence; this assumption is likely too strong to hold across countries in our study, as contexts where IPV prevalence is higher may have lower underreporting due to the normalization of violence. Even within the four countries examined, we observed key differences in the degree of underreporting overall and by type of IPV. The assumption that sensitivity and specificity are independent of prevalence within a given context is more defensible. Therefore, our results should not be generalized to other countries without due consideration.

Implications for Research and Practice

Our findings support best practices in IPV research, which include using multiple behaviorally based items and providing multiple disclosure opportunities.⁶ The substantial underreporting observed in the Albania and Peru DHS, where a single item was used to assess IPV, underscores the need for standardized, multi-item IPV assessment across surveys. While the DHS physical IPV items seem to be performing similarly to violence-focused survey items, efforts are needed to improve the content validity of sexual and emotional IPV by adding new items, such as “having sexual intercourse because she was afraid of him if she refused” and “being intimidated or scared on purpose,” and refining existing items through cognitive interviewing.^{10,43} Items from violence-focused survey used in other countries may also be considered for addition to the DHS measures.⁴⁴ Qualitative studies may elucidate the norms underlying differential stigma attached to various forms of IPV and identify correlates of disclosure.

More rigorous population-based studies comparing anonymous versus face-to-face IPV disclosure could help to expand the use of bias analysis in IPV research. List experiments and sealed envelope methods are promising but need to be done in more contexts and with population-based samples. As such, multiple overimputation or multidimensional bias analysis may only currently be applicable to data from countries in which another high-quality survey has already been conducted in the population of interest to inform priors. For bias adjustment techniques such as multiple overimputation to become more widespread, internal validation studies would ideally become standard practice in large-scale IPV monitoring efforts and in violence prevention trials. In the short term, smaller validation studies could be conducted at the country level to establish reasonable bounds of sensitivity and specificity for IPV scales. Until such procedures become more common, machine learning may be useful for bias adjustment

in violence research where no prior information about underreporting is available.¹³

Finally, this analysis serves as proof of concept to highlight the utility and limitations of two methods for adjusting IPV estimates from DHS and other sources (e.g., medical records, crime data) in countries where a high-quality violence-focused survey or other validation survey can be used to inform priors. Multiple overimputation, which has the advantage of bias-adjusting at the level of individual observations, may be used to reanalyze existing IPV trial data to assess the degree to which measurement error impacted study inference¹⁹ and to analyze new data as more information becomes available about misreporting. While multidimensional bias analysis is an established method in epidemiology, multiple overimputation has been used in only a handful of epidemiologic studies to date.^{35,36} As Bayesian methods and frameworks gain greater prominence in epidemiologic research,⁴⁵ multiple overimputation may be a promising method for bias adjusting self-report data in studies involving other sensitive disclosures such as intravenous drug use, abortion history, and other health behaviors for which clinical verification is difficult.

Conclusion

Multidimensional bias analysis and multiple overimputation provide complementary information about the nature and extent of IPV underreporting: multidimensional bias analysis elucidates reasonable bounds of sensitivity and specificity of measures used in multicomponent surveys, while multiple overimputation is a promising but underutilized method for bias adjustment when there is some information about the level of underreporting from external sources. Multiple overimputation has an advantage over multidimensional bias analysis when it is necessary to bias-adjust individual-level data. Because underreporting appears highly context dependent, our findings should not be applied in countries where there is no prior information about IPV misreporting from other sources. This analysis highlights the need for more research into IPV misreporting, particularly with methods that allow for anonymous disclosure.

REFERENCES

- Sardinha L, Maheu-Giroux M, Stöckl H, Meyer SR, García-Moreno C. Global, regional, and national prevalence estimates of physical or sexual, or both, intimate partner violence against women in 2018. *Lancet*. 2022;399:803–813.
- García-Moreno C, Pallitto C, Devries K, et al. *Global and Regional Estimates of Violence Against Women: Prevalence and Health Effects of Intimate Partner Violence and Non-Partner Sexual Violence*. World Health Organization. 2013.
- Artz S, Jackson MA, Rossiter KR, Nijdam-Jones A, Géczy I, Porteous S. A comprehensive review of the literature on the impact of exposure to intimate partner violence for children and youth. *Int J Child Youth Fam Stud*. 2014;5.
- McTavish JR, MacGregor JC, Wathen CN, MacMillan HL. Children's exposure to intimate partner violence: an overview. *Int Rev Psychiatry*. 2016;28:504–518.
- Bender AK. Ethics, methods, and measures in intimate partner violence research: the current state of the field. *Violence Against Women*. 2017;23:1382–1413.
- United Nations Statistics Office. *Guidelines for Producing Statistics on Violence Against Women: Statistical Surveys*. Vol. 110. UN. 2014.
- Chan D. So why ask me? Are self-report data really that bad? In: *Statistical and Methodological Myths and Urban Legends*. Routledge. 2010;329–356.
- Robinson SR, Ravi K, Voth Schrag RJ. A systematic review of barriers to formal help seeking for adult survivors of IPV in the United States, 2005–2019. *Trauma Violence Abuse*. 2021;22:1279–1295.
- Goodson A, Hayes BE. Help-seeking behaviors of intimate partner violence victims: a cross-national analysis in developing nations. *J Interpers Violence*. 2021;36:NP4705–NP4727.
- Follingstad DR, Bush HM. Measurement of intimate partner violence: a model for developing the gold standard. *Psychol Violence*. 2014;4:369–383.
- Agüero JM, Frisancho V. Measuring violence against women with experimental methods. *Econ Devel Cult Change*. 2022;70:1565–1590.
- Walby S. Improving the statistics on violence against women. *Stat J UN Econ Comm Eur*. 2005;22:193–216.
- Chen Z, Ma W, Li Y, et al. Using machine learning to estimate the incidence rate of intimate partner violence. *Sci Rep*. 2023;13:5533.
- Tanton C, Bhatia A, Pearlman J, Devries K. Increasing disclosure of school-related gender-based violence: lessons from a systematic review of data collection methods and existing survey research. *BMC Public Health*. 2023;23:1012.
- Lépine A, Treibich C, d'Exelle B. Nothing but the truth: consistency and efficiency of the list experiment method for the measurement of sensitive health behaviours. *Soc Sci Med (1982)*. 2020;266:113326.
- Cullen C. Method matters: the underreporting of intimate partner violence. *World Bank Econ Rev*. 2023;37:49–73.
- Gilligan DO, Hidrobo M, Leight J, Tambet H. Using a list experiment to measure intimate partner violence: cautionary evidence from Ethiopia. *Appl Econ Letters*. 2024;32:1594–1600.
- Christopher E, Drame ND, Leyna GH, Killewo J, Bärnighausen T, Rohr JK. Disclosure of intimate partner violence by men and women in Dar es Salaam, Tanzania. *Front Public Health*. 2022;10:928469.
- Bulte E, Lensink R. Women's empowerment and domestic abuse: experimental evidence from Vietnam. *Europ Econ Rev*. 2019;115:172–191.
- Joseph G, Javaid SU, Andres LA, et al. Underreporting of gender-based violence in Kerala, India: an application of the list randomization method. *World Bank Policy Research Working Paper*. 2017(8044).
- Ackerman J. Assessing conflict tactics scale validity by examining intimate partner violence overreporting. *Psychol Violence*. 2018;8:207–217.
- Ackerman JM. Over-reporting intimate partner violence in Australian survey research. *Br J Criminol*. 2016;56:646–667.
- Clark CJ, Shrestha B, Ferguson G, et al. Impact of the change starts at home trial on women's experience of intimate partner violence in Nepal. *SSM Popul Health*. 2020;10:100530.
- Gibbs A, Corboz J, Shafiq M, et al. An individually randomized controlled trial to determine the effectiveness of the Women for Women International Programme in reducing intimate partner violence and strengthening livelihoods amongst women in Afghanistan: trial design, methods and baseline findings. *BMC Public Health*. 2018;18:1–13.
- Heath R, Hidrobo M, Roy S. Cash transfers, polygamy, and intimate partner violence: experimental evidence from Mali. *J Devel Econ*. 2020;143:102410.
- Blackwell M, Honaker J, King G. A unified approach to measurement error and missing data: overview and applications. *Sociol Methods Res*. 2017;46:303–341.
- Fox MP, MacLehose RF, Lash TL. *Applying Quantitative Bias Analysis to Epidemiologic Data*. Springer; 2021.
- García-Moreno C, Jansen HA, Ellsberg M, Heise L, Watts CH; WHO Multi-country Study on Women's Health and Domestic Violence against Women Study Team. Prevalence of intimate partner violence: findings from the WHO multi-country study on women's health and domestic violence. *Lancet*. 2006;368:1260–1269.
- Duvvury N, Marcos M, Gadallah M, et al. *The Egypt Economic Cost of Gender-based Violence Survey (ECGBVS) 2015*. Central Agency for Public Mobilization and Statistics (CAPMAS), Egypt. 2016.
- Ministry of Women's Affairs (Cambodia), National Institute of Statistics (Cambodia), World Health Organization (WHO). *Cambodia WHO*

- Multi-country Study on Women's Health and Domestic Violence Against Women 2015*. National Institute of Statistics (Cambodia).
31. Organization for Security and Co-operation in Europe. *OSCE-Led Survey on Violence Against Women: Main Report*. 2019.
 32. Haarr R. *Violence Against Women and Girls in Albania. National Population Survey*. UNDP Albania, Tiran. 2019.
 33. World Health Organization. *Ethical and Safety Recommendations for Intervention Research on Violence Against Women: Building on Lessons from the WHO Publication Putting Women First: Ethical and Safety Recommendations for Research on Domestic Violence Against Women*. World Health Organization. 2016.
 34. Rutstein SO, Rojas G. *Guide to DHS Statistics*. ORC Macro. 2006;38:78.
 35. Schomaker M, Hogger S, Johnson LF, Hoffmann CJ, Bärnighausen T, Heumann C. Simultaneous treatment of missing data and measurement error in HIV research using multiple overimputation. *Epidemiology*. 2015;26:628–636.
 36. Bengtson AM, Westreich D, Musonda P, et al. Multiple overimputation to address missing data and measurement error: application to HIV treatment during pregnancy and pregnancy outcomes. *Epidemiology*. 2016;27:642–650.
 37. Blackwell M, Honaker J, King G. A unified approach to measurement error and missing data: details and extensions. *Sociol Methods Res*. 2017;46:342–369.
 38. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Vol. 81. John Wiley & Sons. 2004.
 39. Lim BH, Valdez CE, Lilly MM. Making meaning out of interpersonal victimization: the narratives of IPV survivors. *Violence Against Women*. 2015;21:1065–1086.
 40. Abbott P, Teti A. Building decent societies? Economic situation and political cohesion after the Arab uprisings. 2017.
 41. Herrold C, Atia M. Competing rather than collaborating: Egyptian non-governmental organizations in turbulence. In: *Nonprofit Policy Forum*. De Gruyter. 2016;389–407.
 42. Clark CJ, Ferguson G, Shrestha B, et al. Social norms and women's risk of intimate partner violence in Nepal. *Soc Sci Med (1982)*. 2018;202:162–169.
 43. Clark CJ, Bergenfeld I, Cheong YF, et al. Patterns of Women's exposure to psychological violence: A global examination of low-and middle-income countries. *SSM Popul Health*. 2023;24:101500.
 44. Bergenfeld I, Clark CJ, Bengtson AM, Haardoerfer R. Measurement structure and regional invariance of the demographic and health survey intimate partner violence items: a comparative confirmatory factor analysis. *Assessment*. 2025;10731911251340847.
 45. Martinez EZ, Achcar JA. Trends in epidemiology in the 21st century: time to adopt Bayesian methods. *Cad Saude Publica*. 2014;30:703–714.