METHODOLOGICAL NOTE

INNOVATIVE SURVEY TECHNIQUES IN COLLECTING DATA ON VIOLENCE AGAINST WOMEN DURING COVID-19: THE EXAMPLE OF LIST RANDOMIZATION

UN WOMEN | WOMEN COUNT

METHODOLOGICAL NOTE

INNOVATIVE SURVEY TECHNIQUES IN COLLECTING DATA ON VIOLENCE AGAINST WOMEN DURING COVID-19: THE EXAMPLE OF LIST RANDOMIZATION

UN WOMEN | WOMEN COUNT

This methodological note details the survey techniques of remote data collection implemented for the Rapid Gender Assessment Surveys on the Impact of COVID-19 on Violence against Women (VAW RGAs) in 13 countries in 2021. In particular, this note elaborates the list randomization technique and its methodological process from design stage to implementation until statistical analysis. It is informed with strong quantitative and qualitative evidence to support the methodology. This methodological note intends to guide VAW survey implementers in designing effective list randomization survey questions and researchers in producing robust prevalence VAW estimates. Additional resources related to the survey, including statistical global and country reports, survey technical reports and microdata can be found on the UN Women Data Hub.

The authors of this note are Ramya Emandi (Statistics Consultant, Women Count Programme, UN Women), Jessamyn Encarnacion (Inter-Regional Advisor on Gender Statistics, Women Count Programme, UN Women), Sara Gysen (Senior Research Director, Ipsos), Kaitlin Love (Independent Researcher), Papa Seck (Chief, Research and Data section, UN Women), and Rea Jean Tabaco (Statistics Specialist, UN Women).

The authors would like to acknowledge all those who contributed to the execution of the VAW RGAs and the production of this note. Consultations with colleagues from the World Bank initiated the planning of this note. UN Women colleagues from the Ending Violence Against Women (EVAW) Section have been closely involved in the discussion and review process, in particular, Raphaelle Rafin (Programme Specialist), who provided a comprehensive review of the Note. Within the Women Count programme, Lauren Billi (Programme and Knowledge Management Specialist) provided support for the execution of the note and Mika Mansukhani (Communications Specialist) provided production and communication support.

UN Women would also like to thank the Bill and Melinda Gates Foundation for their generous financial support to the VAW RGAs project.

Design: Ouissal Hmazzou
Editing: Jen Ross

EXECUTIVE SUMMARY

Violence against women (VAW) is a human rights violation with often devastating immediate and long-term consequences. Women around the world experience it in various forms, settings, levels of frequency and severity, at the hands of intimate partners, family members or others. Since the start of the COVID-19 pandemic, a wide variety of reports and data have confirmed that the pandemic brought a significant rise in risk factors and intensified VAW for women around the world, at a time when services available to VAW survivors were even more limited.

Limitations in available data and data collection made the process of confirming the hypothesis around the rise in VAW difficult. In 2021, UN Women, with Ipsos as a survey research partner, implemented VAW Rapid Gender Assessment surveys (VAW RGAs) in 13 countries across six regions with support from the Bill and Melinda Gates Foundation, national statistical offices and women's machineries.

Due to the COVID-19 pandemic, this research had to be conducted remotely, via telephone interviews at times when lockdowns and other physical distancing measures were in place in most territories. Additionally, the utmost priority when researching VAW is the safety of respondents and interviewers. To ensure respondents' safety while using remote, non-traditional methods for researching VAW, UN Women explored innovative approaches to asking questions about VAW that would safeguard women's well-being and ensure that the research did no harm. One such technique used by social researchers to elicit truthful responses to sensitive questions is list randomization (LR), which collects data on direct experiences via indirect questioning of respondents that allows them to answer without indicating their specific responses. Benefiting from consultations and guidance of the World Bank and a technical advisory group, UN Women extended the work on the LR methodology to produce intimate partner violence (IPV)[1] estimates through the VAW RGAs even if asked indirectly and remotely.

This note provides guidance on how to reliably estimate the prevalence of IPV that women experience using remote survey methodologies in the context of emergencies or crises, considering both qualitative and quantitative factors. In this case, IPV prevalence estimates both prior to and since the start of the COVID-19 pandemic were calculated. This analysis is presented along with further details on potentially correlated demographic attributes – indicating the characteristics of women who were more likely to experience IPV, both prior to the start of the COVID-19 pandemic and since. Estimates are also provided by country that are statistically significant to reliably report.

Prior to this study, limited remote-based research using LR to understand and report on VAW had been conducted, and there was a lack of empirical evidence on the impact of using LR to validate results and substantiate tailored safety and ethics protocols. Conclusions of this research study support the hypothesis that this is possible, with careful planning and adherence to strong statistical work and safety protocols. Similar to other remote data-collection efforts, implementing LR remotely must include corresponding interviewer training, built-in safety checks within the survey, close monitoring of fieldwork, and detailed observance of verbal and non-verbal cues throughout the interviews.

Ultimately, more studies need be done to build the empirical evidence base supporting the use of LR for VAW, as well as additional research on the use of alternate innovative survey tools to support safe and ethical remote-based data collection. To this end, UN Women calls for more research using these innovative survey tools and methodologies to assess VAW under the necessary condition that the safety of the respondents and interviewers be prioritized. In particular, three aspects related to the use of LR for understanding VAW should be further researched: improving the design and framing of LR questions, guidance for the implementation of LR (including in a face-to-face context), and how to best use LR to measure VAW prevalence to ensure that researchers leave no one behind.

1 In this study, only physical IPV was considered.

HOW LIST RANDOMIZATION WAS IMPLEMENTED IN UN WOMEN'S VAW RAPID GENDER ASSESSMENTS

UN Women's RGAs on the impact of COVID-19 on violence against women

Nearly 1 in 3 women will face physical and/or sexual violence by an intimate partner or sexual violence from a non-partner in her lifetime. It is a human rights violation of pandemic proportions that preceded COVID-19. During the COVID pandemic, administrative data proved to be precious to assess service availability and accessibility. But this alone was insufficient to gauge the magnitude of the impact of COVID-19 on VAW. To inform plans for data collection during lockdowns and restricted mobility periods, UN Women and the World Health Organization (WHO) produced global guidance[2] highlighting important ethical and safety considerations.

The same considerations for women's safety especially hold true for rapid or remote data collection. Given these methodological concerns, UN Women decided to work further on the successful implementation of RGAs in 57 countries that evaluated the impact of COVID-19 on socioeconomic conditions.

Informed by the learnings of this first round of RGAs, in 2021, in order to address the VAW data gaps UN Women implemented VAW RGAs in 13 countries[3] across six regions. This was done with Ipsos as a survey research partner, with financial support from the Bill and Melinda Gates Foundation, and in partnership with national statistical offices (NSOs) and women's machineries. The goal of these surveys was to better understand the extent of this shadow pandemic and produce much-needed data on the current state of women's perceptions and experiences of VAW and safety, both in public and private spaces, as well as their overall mental well-being.

The VAW RGAs were implemented in two phases in Albania, Bangladesh, Cameroon, Colombia, Côte d'Ivoire, Jordan, Kenya, Kyrgyzstan, Morocco, Nigeria, Paraguay, Thailand and Ukraine[4] using a computer-assisted telephone interview (CATI) methodology. Countries were selected based on regional diversity, with priority given to low-middle income countries implementing related UN Women programming Working with a provider that maintains up-to-date telephone databases in each country, mobile phone numbers were randomly dialled, and respondents screened by gender and phone ownership. This approach was deemed the most appropriate given COVID-19 restrictions; however, due to the methodology, respondents were limited to women with access to mobile phones, meaning the sample may be skewed by this factor. Ultimately, data were collected from

2 UN Women and WHO, April 2020, "Violence against women and girls data collection during COVID-19", Available at https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/Library/Publications/2020/VAWG-data-collection-during-COVID-19-compressed.pdf

3 Data collection in 13 countries was undertaken in two phases. Phase I included four countries (specified below) and Phase II included the remaining nine countries. Learnings from Phase I were applied to Phase II, particularly in terms of the list randomization, where significant changes were made to improve methodological soundness and rigour, after consultations with the World Bank. For this reason, only the list randomization results from Phase II are discussed in this paper.

4 Phase I Countries: Cameroon, Kenya, Thailand and Ukraine; Phase II Countries: Albania, Bangladesh, Colombia, Côte d'Ivoire, Jordan, Kyrgyzstan, Morocco, Nigeria and Paraguay.

16,154 women (at least 1,200 per country) ages 18 and older, with quotas set to ensure a nationally representative sample based on geographic and age group distribution.

Cognizant of the fact that COVID-19-related movement restrictions meant women who experienced violence at home would likely be living with their abusers, enhanced ethical and safety protocols were established to ensure that respondents would not be put at any further risk. This included, but was not limited to: additional enhanced training for interviewers to be able to identify when the speakerphone was on, if a recording device might be activated or if anyone was around; rescheduling interviews whenever the respondent was only able to use speakerphone; multiple built-in checks to ensure that the respondent was alone; providing respondents with a safe word that could be used at any time to end the survey; only asking questions that required neutral answers; and providing information about support services to all respondents. This work was guided by a technical advisory group made up of experts in VAW research and programming that have conducted similar work.

Given that directly asking questions on women's experience of violence over the phone may be risky, only indirect questions[5] using innovative techniques such as vignettes, list randomization experimentation, and proxy referencing methods were used to measure VAW.[6] Previous research has suggested that using indirect questions in CATI to measure VAW during COVID-19 has resulted in conclusive evidence.[7]

Vignettes describe an event, happening, circumstance, or other scenario, the wording of which is often experimentally controlled. In VAW RGAs, a hypothetical story with characters whose names were changed depending on the country was narrated to the respondents. For example:

> **"Mary and John are a couple. They have been married for several years and have two children. John works in a repair shop, but lately the business has been bad, and they are worried about money. Sometimes when John gets stressed, he takes out his anger by yelling at Mary, and sometimes he hits her. Mary feels hurt and wants him to stop but does not know what to do."**

Respondents were then asked whether this is common or not in the area where they live. This serves as an indirect question to collect data about norms on VAW in a community.

Indirect proxy questions that measure perceptions around the community-level occurrence (or increase) in violence were also used in VAW RGAs. Respondents were asked whether they think violence against women is a problem in the area where they live and whether it has increased since COVID-19.

In list randomization, respondents are asked to answer questions based on their own experience with a sensitive topic asked alongside other non-sensitive experiences – a strategy commonly used when conducting research about a variety of potentially sensitive topics that could invoke social desirability bias, for instance politics, corruption, abortion, illegal migration, racial discrimination, among others (see an example on in Section 3).

5 The exception to this was in Colombia, where the technical advisory group and research team felt confident that direct questions about violence could be asked without compromising the safety of respondents.

6 For the purpose of the assessments, experiences of VAW were defined as: physical abuse (i.e., been slapped, hit, kicked, had things thrown at them, or other physical harm); verbal abuse (i.e., being yelled at, called names, humiliated); denied basic needs (i.e., health care, money, food, water, shelter); denied communication (i.e., with other people, including being forced to stay alone for long periods of time); and sexual harassment (i.e., being subjected to inappropriate jokes, suggestive comments, leering or unwelcome touch/kisses).

7 Peterman, Amber. 2021. "The Art of Indirect Measures: Asking about Violence Against Women and Children in Remote Surveys." Center for Global Development.

When it comes to question design and statistical analysis, both vignettes and proxy referencing are similar to regular survey questions measuring perceptions and attitudes, but neither can generate prevalence estimates. Only list randomization can generate prevalence estimates, but it needs an intricate design as well as statistical rigour in its analysis. Thus, guidance on the use of list randomization is needed to carefully design, implement and analyse LR results.

Introduction to list randomization

Over time, social researchers have developed a number of techniques to mitigate social desirability bias by encouraging truthful responses to sensitive questions. One such method of asking sensitive questions indirectly is list randomization – also known as the item count or unmatched count technique – that provides a desensitized approach for respondents to report on sensitive behaviour without indicating to the interviewer their personal response.

To apply this technique, survey respondents are randomly split into two groups (control and treatment) who are each presented with a short list of statements, typically three or four. Survey respondents in the treatment group receive the same list of statements and one key additional statement designed to capture sensitive information. After being presented with all statements, respondents are asked to indicate how many statements are true, without indicating which specific statements they believe are true within the list. By subtracting the mean number of true statements in the treatment group from the mean number of true statements in the control group, it is possible to estimate the proportion of the sample that has been exposed to the sensitive behaviour. Several studies have suggested that this technique can yield more accurate responses to sensitive questions when compared to direct reporting.[8]

UN Women and Ipsos benefited tremendously from technical assistance from the World Bank on the development of list randomization techniques for estimating VAW.[9] Significant changes were made to the list randomization exercise between Phase I and Phase II of the research, and therefore Phase II findings will be primarily discussed in this paper.

There is significant research that supports the use of list randomization as an effective way to mitigate risks associated with self-reported behaviour and beliefs, topics that have been a challenge for survey researchers across disciplines for decades. This methodological note has highly benefited from this literature, which is discussed throughout the note.

8 Across 48 comparisons of direct report and list randomization, one meta-analysis found that 63% of estimates for socially undesirable behaviour were significantly larger when elicited through list randomization (Holbrook and Krosnick 2009). A more limited meta-analysis found that while list randomization estimates of socially undesirable behaviour were generally larger, particularly for studies using undergraduate samples, the overall difference was not significant (Tourangeau 2007).

9 Diego Javier Ubfal, Elizaveta Perova, Ervin Dervisevic, from the World Bank were consulted. The work developed by the World Bank's East Asia and Pacific Gender Innovation Lab in 2020 on "Can we capture exposure to Gender-based Violence (GBV) through Phone Surveys during a Pandemic?" was used as a reference for UN Women's VAW RGAs work.

HOW TO DESIGN, IMPLEMENT AND ANALYSE LR

The VAW RGAs aimed to measure IPV over two time periods: before and after the onset of COVID-19. To capture women's experiences, list randomization was implemented using two sets of questions. The control group was presented with the following questions with non-sensitive statements (control items):

Question 1

Please tell me how many of the following statements you regard as true:

1. I prefer [LOCAL FOOD ONE] to [LOCAL FOOD TWO]
2. I like [NAME OF WELL-KNOWN LOCAL MUSICIAN, BUT NOT ONE EVERYONE LIKES]
3. Women in my family enjoy watching [POPULAR LOCAL SPORT].

Question 2

Please tell me how many of the following statements you regard as true:

1. People in my family have played [POPULAR LOCAL SPORT].
2. I like [NAME OF WELL KNOWN TV SHOW, BUT NOT ONE EVERYONE LIKES]
3. I prefer [LOCAL FOOD THREE] to [LOCAL FOOD FOUR]

For the treatment group, a sensitive item (shown as statement 4 below) was added to the same list used for the control group.

Question 1

Please tell me how many of the following statements you regard as true:

1. I prefer [LOCAL FOOD ONE] to [LOCAL FOOD TWO]
2. I like [NAME OF WELL KNOWN LOCAL MUSICIAN, BUT NOT ONE EVERYONE LIKES]
3. Women in my family enjoy watching [POPULAR LOCAL SPORT].
4. I have been slapped or hit by my husband or partner BEFORE THE ONSET OF COVID-19

Question 2

Please tell me how many of the following statements you regard as true:

1. People in my family have played [POPULAR LOCAL SPORT].
2. I like [NAME OF WELL KNOWN TV SHOW, BUT NOT ONE EVERYONE LIKES]
3. I prefer [LOCAL FOOD THREE] to [LOCAL FOOD FOUR]
4. I have been slapped or hit by my husband or partner AFTER THE ONSET OF COVID-19

The order of the statements was randomized automatically for each respondent by the CATI programming. The non-sensitive statements were carefully selected in consultation with technical experts from the World Bank, and the control items in both list experiments fit as non-sensitive, common items in all local settings. The design validity of the list experiments is based on assumptions that are discussed in the next subsection.

Other considerations implemented to successfully design, implement and analyse LR are outlined in this section.

Sampling: Ensuring randomization and representation of control and treatment groups

During the sample design, survey respondents must be divided randomly in two groups: control and treatment groups. Randomization ensures that each respondent has an equal chance of receiving any of the treatments under study (here, in this case, it is the exposure to the sensitive statement), and generate comparable intervention groups. This ensures that the treatment is the only source of potential differences in outcomes between the two groups. Table 1 presents how the samples were distributed in the VAW RGA countries.

TABLE 1.
Sample size of treatment and control groups in VAW RGAs, by country*

Country	Treatment Group	Control Group
Albania	604	606
Bangladesh	616	626
Colombia	609	600
Côte d'Ivoire	640	685
Jordan	603	601
Kyrgyzstan	602	599
Morocco	609	605
Nigeria	770	712
Paraguay	622	588

*Unequally sized groups are common in research and may be the result of sampling weights, simple randomization and/or study dropouts.

To ensure representation in each group, it is often advised to use block sampling by demographics that will be analysed, such as age, gender and urban/rural location, etc. Not limited to demographic indicators, block sampling may include indicators that capture risk factors of VAW or that record respondent's history of VAW, if data are available. Using block sampling rather than simple randomization is preferable to achieve balance across participant attributes that researchers are later interested in analysing.

Design: Formulating list randomization questions and statements to minimize design effects

Avoid correlation among the control statements

One of the major considerations while conducting LR is to make sure that the sensitive and non-sensitive statements within the list are not correlated in any way. However, this was not the case in the set of statements designed during the Phase I of the survey. Between Phase I and Phase II, the non-sensitive statements included in the list randomization were changed to be more neutral and less likely to be correlated with the outcome the list randomization was intended to capture. For instance, initially, statements included: "I have faced severe economic hardships in my life" and "I feel depressed on most days". These statements may initially seem innocuous, however experience of intimate partner violence (the outcome that was to be measured) is correlated with socioeconomic status and mental health outcomes. Non-correlation can be assured by asking neutral statements for control items such as "having seen a movie", "to talk to someone local", etc. or even to have pop culture reference statements. The care taken in ensuring no correlation among the statements can reduce design effects. Accordingly, the statements were readjusted prior to Phase II of this research to ensure non-correlation among them and are presented in Annex 1. Additionally, the sensitive statements were also designed to be the same for tracking experiences both before and after COVID-19 in order to avoid differentiated bias between statement and to measure how estimated prevalence differs before and after the onset of the pandemic.

Conduct cognitive testing

It is a best practice to observe respondent understanding and reactions to the list randomization questions in order to gauge whether respondents really understood the techniques used in the survey. In the case of this research study, this was done in two ways.

First, cognitive testing was conducted in-person in select countries. The list randomization exercise was found to keep participants engaged and facilitate expression. Participants found the exercise easy to follow, and they said it provided them with a way to more comfortably and indirectly disclose violence they may have experienced. Despite the sensitivity inherent in asking questions about women's experiences with physical abuse, it was deemed important by all participants who appreciated the opportunity to express themselves and raise awareness on IPV.[10]

Feedback provided by interviewers during debriefing sessions indicated that the vast majority of respondents who participated in the survey welcomed the opportunity to discuss their experiences and make their voices heard. This is not to undermine the difficult experiences of violence that women may have faced, given personal histories and trauma.

Provide detailed instructions to guide respondents and gauge their reactions

Supplementary qualitative measures, such as including instructions for respondents, can bring clarity to the question and gather better responses. For the LR questions specifically, a question was added to assess the interviewer's perception of respondent's understanding of the LR questions with an additional question posed only to the interviewer in the survey. Interviewers were asked to assess whether they think the respondent understood how to answer the LR question (for instance, did the respondent ask for instructions to be repeated? / take a long time to answer? / seem unsure?).

Based on field observations of the RGA pilot survey, the RGA included an instruction after the list randomization statements asking respondents to use their fingers to count the statements they regarded as true. This aimed to avert their repetition of any of the statements aloud, so no others in the household could overhead them mention these items. Providing only four statements likely made this exercise easier. An increase in the number of statements could make the exercise more difficult, particularly for those who are less numerate. Enumerators reported that this worked well for participants to keep track of their responses.[11]

Analyse dropout rates for LR questions

During this research, 2% of respondents said they did not want to participate in the survey during the screening, perhaps indicating that they did not feel safe in doing so. Among those who continued, only 15% dropped out during the course of the survey, indicating that the vast majority of participants felt comfortable with the topics discussed. The dropout at LR questions was also very low across all countries and these questions were observed by interviewers to be a bit more difficult to understand only in Albania, Colombia and Thailand.

While the dropout rates at the LR section of the survey were overall very low, they did differ based on whether the respondent received the LR that included the sensitive statement. At the first neutral LR question, the drop-out rate among respondents in the control group was 0.07%, compared to 0.15% in the treatment group that received the sensitive statement. Similarly, at the second LR question the drop-out rate among those in the control group was 0.04%, compared with 0.12% in the treatment group that received the sensitive statement. While these numbers are very low, representing 47 respondents in a study of over 10,000, it is important to note the slight difference in drop-out rates between the two groups.

10 UN Women. 2021. Cognitive Testing Report: Rapid gender assessment on the impact of COVID-19 on violence against women.

11 Porter, Catherine, Marta Favara, Alan Sánchez and Douglas Scott. 2021. "The impact of COVID-19 lockdowns on physical domestic violence: Evidence from a list randomization experiment," Elsevier SSM – Population Health.

Piloting: Checking for floor and ceiling effects to minimize the possibility of design effects

Floor and ceiling effects represent two respondent behaviours that may interfere with the ability of list experiments to elicit truthful answers. Ceiling effects may result when respondents' true preferences are affirmative for all the control items as well as the sensitive item. Floor effects may arise if the control questions are so uncontroversial that uniformly negative responses are expected for many respondents. Another possible floor effect may arise if respondents fear that answering "0" reveals their truthful (negative) preference. Under both scenarios, respondents in the treatment group may fear that answering the question truthfully would reveal their true preference for the sensitive item.[12]

Checking for the lack of ceiling and floor effects in the pilot phase gives the opportunity to select non-sensitive control statements that minimize the likelihood of potential responses that trigger such effects. Piloting enables the careful selection of control items as a crucial part of the experiment to reduce design effects.

Post-data collection: Statistical testing for internal validation of data prior to the analysis

The validity of the LR methodology can be established by testing three assumptions:

1. successful randomization of the treatment
2. the absence of ceiling and floor effects
3. no design effects[13]

Successful randomization is required such that individuals allocated to each group are, on average, likely to agree with the same number of non-sensitive statements in any given list. The absence of ceiling and floor effects is required, as individuals may be reluctant to provide truthful answers if they believe they no longer benefit from the privacy of their responses. Finally, the 'no design effects' assumption is necessary so that the inclusion of the sensitive item does not change the number of positive answers to the non-sensitive items.

Test for randomization:

The success of the treatment randomization was assessed by comparing the means of a series of individual and household characteristics among the treated and control groups. Successful randomization of the groups is assessed by computing the t-test which is to compare the means of the treatment and control groups. Consider the null hypothesis and calculate the t statistic based on the sample data.

Hypotheses:

H_0: $\mu_1 = \mu_2$. The null hypothesis is the means of when the two groups are equal.

H_A: $\mu_1 \neq \mu_2$. The alternate hypothesis is the means of when the two groups are not equal.

In Table 2 results of the t-test provide sample means for the main variables in the two treatment groups. T-test results for all categories have p-values > 0.05; hence the researchers fail to reject the null hypothesis H_0 and the means of the samples are statistically equal in both the groups. Comparing the sample means between treatment and control suggests that the randomization of the list experiment was successful, given that all observable characteristics do not significantly differ between the two groups.

12 Kuklinski, J. H., M. D. Cobb, and M. Gilens. 1997a. "Racial attitudes and the 'New South'." Journal of Politics. 59. pp. 323–49; Kuklinski, J. H., P. M. Sniderman, K. Knight, T. Piazza, P. E. Tetlock, G. R. Lawrence, and B. Mellers. 1997b. "Racial prejudice and attitudes toward affirmative action." American Journal of Political Science. 41. pp. 402–19.

13 Lépine, A., Treibich, C., & D'Exelle, B. 2020a. "Nothing but the truth: Consistency and efficiency of the list experiment method for the measurement of sensitive health behaviours." Social Science & Medicine. 266: 113326.

TABLE 2.
Mean and p-value of main variables

	Treatment group	Control group	t-test
Sociodemographic characteristics	(mean)	(mean)	(p-value)
Marital status			
Marital status (living with partner)	0.61	0.60	0.340
Marital status (separated/divorced)	0.04	0.04	0.368
Marital status (single)	0.29	0.29	0.954
Marital status (widowed)	0.06	0.06	0.315
Is disabled	0.19	0.19	0.957
Is employed	0.40	0.39	0.513
Generating income	0.12	0.12	0.956
Has children	0.70	0.69	0.389
No. of children	1.61	1.60	0.872
Education status			
No formal education	0.11	0.11	0.478
Completed primary education	0.23	0.23	0.319
Completed secondary education	0.29	0.28	0.473
Completed college/university	0.25	0.27	0.241
Completed post graduate studies	0.03	0.03	0.986
Urban	0.78	0.79	0.721
Rural	0.22	0.21	0.721
Living in households with loss of income	0.68	0.69	0.769
No. of household members	5.05	5.11	0.344
Age groups			
Age group 1 (18 – 29 years)	0.34	0.35	0.565
Age group 2 (30 – 39 years)	0.26	0.25	0.228
Age group 3 (40 – 49 years)	0.17	0.17	0.804
Age group 4 (50 – 59 years)	0.12	0.12	0.517
Age group 5 (60+ years)	0.10	0.10	0.832
Observations	4,345	4,306	

Testing for the absence of ceiling and floor effects:

Respondents in the treatment group may fear that answering the question truthfully would reveal their true preference for the sensitive item. Ceiling effects occur may when the respondent feels that all statements are true. In this case, she may not want to report a count of '4 truths' as she might fear losing her own desired level of protection and being identified as a victim of physical violence.

On the contrary, floor effects occur when the response is that none of the statements are true. If the respondent should only feel that one statement is true, she may not report a count of 1 truth – and instead report 0 – to avoid running the risk of revealing that her response is the targeted statement on violence against women.

Table 3 presents the response rate for all possible answers by respondents in both treatment and control groups. The results suggest that the responses to 0 statements true and all 4 statements true is not abysmally low for both the list experiments; hence, no significant ceiling and floor effects are observed.

TABLE 3.
Percentage distribution and mean of responses in list randomization questions

List Randomization Experiment 1

Before the onset of COVID-19	Distribution of respondents by number of statements that are true (per cent)					Mean of all responses
	0	1	2	3	4	
Treatment group	10.6	43.0	33.7	10.2	2.4	1.54
Control group	12.6	46.5	31.5	9.4	-	1.43

List Randomization Experiment 2

After the onset of COVID-19	Distribution of respondents by number of statements that are true (per cent)					Mean of all responses
	0	1	2	3	4	
Treatment group	7.4	37.6	32.5	16.1	6.2	1.74
Control group	7.4	37.2	38.0	17.3	-	1.61

Testing for design effects (No liars test):

The final and perhaps most important assumption of the list experiment is the absence of design effects, which occurs when the inclusion of the sensitive item affects respondents' answers to non-sensitive items. Based on the inclusion of the sensitive statement, the respondent might change their responses to control statements. Blair and Imai formulated a null hypothesis test to evaluate the design effects.[14] The null hypothesis of the test indicates no design effects and can be represented by the following:

$$Pr(Y_i \leq y \mid T_i = 0) \geq Pr(Y_i \leq y \mid T_i = 1)$$
$$\forall\, y = 0,\ldots,3$$

$$Pr(Y_i \leq y \mid T_i = 1) \geq Pr(Y_i \leq y - 1 \mid T_i = 0)$$
$$\forall\, y = 1,\ldots,4$$

where $T_i = 1$ ($T_i = 0$) implies that the respondent i belongs to the treatment (control) group.

Here, the potential outcome of a respondent to a control statement is tested such that when in the treatment group and when in the control group, the respondent should have the same response. In other words, the proportion of individuals in the control group who agree with no more than y statements (y =0, 1, 2, 3) should be greater than this proportion for the treated group, and the latter proportion (for y = 1, 2, 3, 4) should be greater than the proportion of individuals in the control group who agree with no more than y - 1 statements. If this rationale is not the case, given that individuals in the treated and control groups are similar on average, it means that individuals in the treated group modified their answers to the nonsensitive items.

Equivalently, the null hypothesis implies that $\pi_{yt} \geq 0$ for all y and T, where

$$\pi_{y1} = Pr(Y_i \leq y \mid T_i = 0) - Pr(Y_i \leq y \mid T_i = 1)$$
$$\forall\, y = 0,\ldots,3$$

$$\pi_{y0} = Pr(Y_i \leq y \mid T_i = 1) - Pr(Y_i \leq y - 1 \mid T_i = 0)$$
$$\forall\, y = 1,\ldots,4$$

If all of the π_{yt} are nonnegative, the null hypothesis of no design effect cannot be rejected; conversely, if all are negative, then the H_0 is rejected. If at least one is negative, it is important to understand if it is negative by chance, using the Bonferroni-corrected p-values.[15] As Table 4 suggests, we check for Bonferroni corrected p-value as one or more values (π_{yt}) are negative and the values of Bonferroni are greater than 0.05 and therefore, we fail to reject the null hypothesis. i.e. there are no design effects.[16]

14 Blair, Graeme and Kosuke Imai. 2012. "Statistical Analysis of List Experiments," Political Analysis. 20: pp. 47 77.

15 The principal idea of the testing procedure is first to conduct a separate hypothesis test for each of the two stochastic dominance relationships given in the null hypothesis, and then combine the results using the Bonferroni correction (Blair and Imai 2012).

16 If this value is below alpha, the null hypothesis of no design effect is rejected. If it is above alpha, null cannot be rejected.

TABLE 4.
Bonferroni-corrected p-value for list experiments to check for design effects

List Experiment 1				
Before the onset of COVID-19	0	1	2	3
	-0.0020	0.0483	0.0279	0.0350
	0.1109	0.4246	0.2733	0.0821
Bonferroni-corrected p-value	0.76460			

List Experiment 2				
After the onset of COVID-19	0	1	2	3
	-0.0083	0.0227	0.0540	0.0642
	0.0831	0.3786	0.3062	0.0995
Bonferroni-corrected p-value	0.152036			

The assumption of design effects for all nine countries under Phase II of the survey does not hold true. When checked for country-specific results, only Kyrgyzstan and Colombia did not meet the no design effects criteria. Hence, the data set for modelling excludes these countries for analysis and comprises seven countries in the final data set (ie. Albania, Bangladesh, Côte d'Ivoire, Jordan, Morocco, Nigeria and Paraguay). Please refer to Annex 2.

COMPUTING FOR PREVALENCE OF IPV USING LIST RANDOMIZATION

There are two ways of to estimate prevalence: a) without covariates, which is estimating the relationship of a respondent being in the treatment group and their prevalence of IPV devoid of individual's characteristics; and b) with covariates, which includes characteristics of respondents in the model (such as age, poverty, employment status, etc.) that may influence the prevalence of IPV.

Estimating without covariates is essentially using the difference-in-means estimator method that is represented mathematically below. The average sensitive behaviour prevalence rate is then given by δ and corresponds to the average difference between the number of statements that the control group and the treatment group agreed with for the list.

To estimate the prevalence of sensitive behaviour:

$$Y_i = \lambda + \delta T_i + \varepsilon_i$$

Y_i =outcome of interest

indicates the number of true responses by individual i to the listed statements

δ= the average sensitive behaviour prevalence rate corresponds to the average difference between the number of statements that the control group and the treatment group agreed with.

T_i =Treatment variable (takes 0 for control group and 1 for treament group)

ε_i=error term

The limitation of using the difference-in-means method without covariates is that it does not allow an efficient estimation of relationships between preferences for the sensitive item and respondents' characteristics. The second way of estimation is with control variables (covariates) – i.e. the respondent's sociodemographic characteristics, which may be correlated with increased violence against women. The covariates chosen are mentioned in Annex 3. For the purpose of the regression model and its interpretation, the categories are classified into binary categories, such as women of reproductive age and not, women with formal education and not, etc.

Among the various estimation strategies, including maximum likelihood and non-linear regression, the researchers choose linear regression in the model identified to produce robust results. "They ran regressions, conditional on individuals' sociodemographic characteristics:

$$Y_i = \alpha + \beta X_i + \delta T_i + \gamma X_i T_i + \varepsilon_i$$

Y_i=outcome of interest,

indicates the number of true responses by individual i to the listed statements

X_i= individual and household characteristics

δ=estimated coefficient,measures the increase in the number of true responses

resulting from receiving the additional sensitive statement

T_i=Treatment variable (takes 0 for control group and 1 for treament group)

γ= interaction term,estimated prevalence for subgroup

ε_i=error term

The researchers were interested in δ and γ, which inform the prevalence estimates. δ measures the increase in the number of true responses due to the additional sensitive statement, while γ measures the prevalence of the interaction term – that of the subgroup of an individual based on the individual characteristic in the treatment group.

IPV Prevalence estimates

The findings are presented in Table 5 for both methods – difference-in-means (without covariates) and linear regression (with covariates). The pooled IPV prevalence estimates for the aforementioned seven countries before the onset of COVID-19 was estimated at 13.2% considering the methodology without covariates and 12.6% considering the methodology with covariates. The results are strongly significant and robust in view of standard errors and the p-value. Similarly, the prevalence estimate for IPV since the onset of COVID-19 is 10.7% (without covariates) and 10% (with covariates).

TABLE 5.
IPV prevalence estimates with and without covariates

	Without covariates	With covariates (Control variables)
IPV before the onset of COVID-19	0.13240*** (0.01857)	0.126106*** (0.018328)
Prevalence estimate	13.2%	12.6%
IPV since the onset of COVID-19	0.10676*** (0.0201)	0.09957*** (0.01966)
Prevalence estimate	10.7%	10.0%

Note: Standard errors are reported in parentheses * p < 0.1, **p <0.05, ***p < 0.01 It is important to note, that the VAW RGA results should not be interpreted as indicating a decrease in physical IPV across the surveyed countries. Analysis must consider the methodological variables, including the differences in timeframe. "Before the onset of COVID-19" spreads over respondents' lifetime, while "IPV after the onset of COVID-19" captured a maximum period of 18 months of experience at the time of data collection; thus, the exposure to risk is higher in the former than the latter. Further, comparisons cannot be made with neither global estimates[17] nor other available prevalence data from countries mainly due to the difference in the reference population; that is, national prevalence surveys ask ever-partnered women mostly of age 15-49 or 15 and older on IPV, while the VAW RGAs asked all women respondents aged 18 or over.

Based on significant and robust results shown in Table 6 below, specific groups of women are more likely to experience IPV before COVID-19:

- Women living in urban areas (16% more likely than women living in rural areas)
- Married women (4% more likely than non-married women)
- Women with child/ren (9% more likely than women without a child)
- Employed women (21% more likely than non-employed women)
- Women earning an income (17% more likely than non-earning women)
- Women from households with decreased income (5% more likely than women from households without decreased income)

Similarly, based on significant and robust results, the specific groups of women who are more likely to experience IPV since COVID-19 are:

- Women of reproductive age (8% more likely than women over 50)
- Women living in urban areas (16% more likely than women living in rural areas)
- Women with child/ren (6% more likely than women without a child)
- Employed women (29% more likely than non-employed women)
- Women earning an income (27% more likely than non-earning women)
- Women without a disability (12% more likely than women with a disability)

17 As of 13 March 2023, global estimates for physical IPV are not yet available.

TABLE 6.
IPV prevalence using covariates

Dependent variable	IPV before the onset of COVID-19	IPV since the onset of COVID-19
Predicted response to sensitive item	0.126106***	0.09957***
	(0.018328)	(0.01966)
Reproductive age	-0.010834	0.08866***
	(0.023985)	(0.02573)
No formal education	0.037730	-0.06459
	(0.030735)	(0.03297)
Urban areas	0.160403***	0.15776***
	(0.020296)	(0.02177)
Married	0.038729*	-0.03437
	(0.019562)	(0.02098)
Have a child	0.091009***	0.05923*
	(0.02153)	(0.02310)
Employed	0.213064***	0.29101***
	(0.020627)	(0.02213)
Earning an income	0.173814***	0.27326***
	(0.029997)	(0.03218)
Disabled	-0.001087	-0.12251***
	(0.026929)	(0.02889)
HH with decreased income due to the pandemic	0.048496*	0.01324
	(0.020646)	(0.02215)
Constant	1.053***	1.341***
	(0.037)	(0.040)
Observations	8,651	8,651
R2	0.034	0.053
Adjusted R2	0.033	0.051
Residual Std. Error (df = 8641)	0.913	0.979
F Statistic (df = 9; 8641)	30.543***	49.959***

Note: standard errors are reported in parentheses * p < 0.1, **p <0.05, ***p < 0.01.

Based on the variables that are significantly affected, both in the periods before and after the pandemic, the model is built on the common control variables. The common variables are urban area, employed, earning an income and having a child. The significant results from this regression model in Table 7 suggest that 9% of women with children are estimated to have experienced IPV before the COVID-19, and that 11% of employed women and 15% of women with an income are estimated to have experienced IPV since the COVID-19.

TABLE 7.
IPV prevalence using common control covariates

Dependent variable	IPV before the onset of COVID-19	IPV since the onset of COVID-19
Predicted response to sensitive item	0.124*** (0.046)	0.096* (0.050)
Urban areas	0.175*** (0.028)	0.201*** (0.030)
Have a child	0.057* (0.030)	0.071** (0.032)
Employed	0.233*** (0.029)	0.260*** (0.031)
Earning an income	0.212*** (0.042)	0.224*** (0.045)
Treat*urban areas	-0.045 (0.040)	-0.065 (0.043)
Treat*having a child	0.086** (0.042)	-0.017 (0.045)
Treat*employed	-0.057 (0.041)	0.109** (0.044)
Treat*earning an income	-0.098* (0.059)	0.146** (0.064)
Constant	1.113*** (0.033)	1.350*** (0.035)
Observations	8,651	8,651
R2	0.034	0.047
Adjusted R2	0.033	0.046
Residual Std. Error (df = 8641)	0.913	0.982
F Statistic (df = 9; 8641)	33.667***	47.528***

Note: standard errors are reported in parentheses * p < 0.1, **p <0.05, ***p < 0.01.

Endogeneity refers to the possibility that the chosen control variables correlate with the error term of the regression model. For instance, the covariate 'Women Being Employed' can relate to families or partners that support women to be independent and progressive. This may in turn relate to partners that do not engage in IPV. Hence, the researchers tested for possible covariates that might cause endogeneity using the Hausman test (see Annex 4 for more information) and found no endogeneity.

Correspondingly, country prevalence estimates were evaluated. The significant results for the seven countries in the data set are shown in Table 8 below (see Annex 5 for the detailed regression results per country):

TABLE 8.
Prevalence estimates of IPV before and after the onset of COVID-19, by country

Countries	Prevalence estimates	
	IPV before the onset of COVID-19	IPV after the onset of COVID-19
Albania	No significant results	No significant results
Bangladesh	13.5%	No significant results
Côte d'Ivoire	14.0%	23.2%
Jordan	No significant results	12.0%
Morocco	33.0%	24.3%
Nigeria	No significant results	16.3%
Paraguay	No significant results	No significant results

Options of external validity

Other options to correspond the results for external validity can be to assess the results with the available statistics. This is, however, not recommended for many factors, such as when the reference period of the two estimates do not match, when the geographical coverage does not match, when the type of prevalence considered does not match, when modes of survey techniques do not match, etc. However, if there is an alternate data set available that is assessed for same type of violence in the same reference period, considering a similar geography and survey implementation, then the prevalence estimates can be compared. Some options that can be considered for comparison are, given the similarities in surveys/data sets as discussed above:

- The prevalence estimates can be compared to other similar related questions within the same survey vis-à-vis list randomization results.
- It is a good practice to include a direct ques-tion on the same question asked in the list randomization question within the same survey, provided the safety considerations of the respondents are not compromised. The results can then be compared between the direct questions and the list randomization estimates. This computes the social desirability bias.
- The prevalence estimates can be compared with the results of LR question on IPV with the latest official statistics or other surveys (such as Demographic and Health Surveys), provided that the surveys are similar in nature, as discussed above.

CONCLUSION AND RECOMMENDATIONS

LR can provide an estimate of VAW prevalence

List randomization experiments supplement to the direct survey measure, and they may provide a more robust estimate of IPV, particularly given known underreporting of this experience. However, it is important to only implement LR if the following statistical and methodological pretexts are conducted:

- Ensure randomization and representation between and within treatment and control groups.
- Ensure that control statements used in LR are not directly OR indirectly correlated with the sensitive statement.
- Train interviewers on conducting the LR experiment and listen for indications that the respondent may be uncomfortable or otherwise not understand the question.
- Check for floor and ceiling effects in order to minimize the possibility of design effects at the analysis phase.
- Use statistical testing to internally validate data prior to conducting any analysis.

List randomization data can be remotely collected in an ethical and safe manner

Data collection using list randomization experiments is a strong tool to increase the safety of remote data-collection methods. This is complemented with the results of qualitative assessments. Not only were the item non-response rates (below 1% in most countries) and drop-out rates (below 1% in all countries) for the LR questions with the sensitive statement very low, but the interviewers also did not report any other issues with these questions. Only in Albania, Colombia and Thailand were the questions found to be a bit more difficult to understand. The interviewers handled these situations by offering additional explanation and repeating the instructions to the respondents.

The authors recommend the following guidelines for ethical and safety instructions to adhere to while implementing LR experiments. This should be complemented with other safety protocols on VAW, refer to WHO guidance[18] and, specifically for CATI, refer to UN Women's evidence-based technical guidelines on collecting violence against women data through telephone interviews.[19]

- List randomization is hard to implement and takes time. In planning for the survey and estimating the duration, include time for participants to understand the exercise.
- Carefully design LR questions by providing detailed instructions.
- Conduct cognitive testing of LR questions to assess the respondent's understanding and conduct a pilot such that the correlation with the statements used in the experiment can be tested.
- When using list randomization, avoid a large number of items/statements to prevent participant fatigue.

Overall, responses for list randomization experiments are based on recall and might trigger survey recall error. Respondents need to answer the survey questions based on the reference period and their memory. For the pre- and post-COVID-19 experiences of IPV, the survey chose to ask two questions in one wave itself. Other methodologies, such as difference-in-difference methods, produce similar

18 https://apps.who.int/iris/bitstream/handle/10665/65893/WHO_FCH_GWH_01.1.pdf?sequence=1&isAllowed=y
19 https://data.unwomen.org/sites/default/files/documents/Publications/Guidance_VAW_RGA-EN.pdf

estimates but the survey must be conducted in two waves with two reference periods. The methodology allows for conducting the survey in one wave.

More studies should be done to build the empirical evidence on LR for VAW and research on the use of alternate innovative survey tools

The authors call for more research to test innovative survey tools and methodologies to assess VAW, provided that the survey implementation takes utmost priority in protecting the safety of the respondents, especially women. LR should be further researched for:

- Improving statements with a list experiment
- Implementation of LR experiments (and maybe included in face-to-face surveys)
- Measuring prevalence estimates, disaggregated by women groups, to ensure that researchers leave no one behind.

ANNEXES

Annex 1

Adjustments in Phase I questions, due to correlation issues

LR statements	Points of correlation/problems in capturing aspects of violence	Proposed change
First LR (C28)		
1. I have faced severe economic hardships in life.	Economic hardship and VAW are correlated.	I prefer [LOCAL FOOD ONE] over [LOCAL FOOD TWO].
2. I want to open a new business.	Risk-taking and confidence are related, and confidence is impacted by VAW as seen in study.	I like [INSERT NAME OF WELL-KNOWN LOCAL MUSICIAN].
3. I have been slapped by my husband or partner.	-	I have been slapped or hit by my husband or partner BEFORE THE ONSET OF COVID-19.
4. I want to have children/more children.	Those with abusive partners may not want more children; VAW and household stress are impacted by children.	People in my family enjoy watching football.
Second LR (C29)		
1. I am jealous of my neighbour.	Those with abusive partners may indeed be jealous of relationships of their neighbours if they are experiencing abuse.	People in my family have played football.
2. My husband/partner has insulted me in front of others.	This statement is double-barrelled and combines VAW in public/private spheres.	I have been slapped or hit by my husband or partner AFTER THE ONSET OF COVID-19.
3. My friends respect me.	Those with abusive partners may indeed feel like their friends do not respect them because of their abusive relationship.	I like [POPULAR LOCAL TV SHOW].
4. I feel depressed most days.	Those with abusive partners may indeed experience depression because of their abusive relationship.	I prefer [LOCAL FOOD TWO] over [LOCAL FOOD ONE].

Annex 2

Testing design effects, by country

Countries	Bonferroni-corrected p-value	
	IPV before the onset of COVID-19	IPV after the onset of COVID-19
All Phase II countries (9)	0.007	0.002
Kyrgyzstan	0.004	0.007
Colombia	0.000	0.022
Bangladesh	1.000	0.814
Côte d'Ivoire	1.000	0.731
Paraguay	0.114	0.183
Nigeria	1.000	0.625
Jordan	0.041	0.247
Morocco	1.000	1.000
Albania	0.082	0.962
Phase II, except Kyrgyzstan and Colombia (7)	0.765	0.152

Annex 3

List of covariates used in estimating the prevalence of sensitive behaviour

Characteristics (covariates)	Type	Categories	Conversion to dummy variables
Age category	categorical	18–29; 30–39; 40–49; 50–59; 60+	Women in reproductive age vs. women in non-reproductive age
Education	categorical	Less than primary/primary; Completed secondary; Technical; College or advanced education	Women with no formal education vs. women with formal education
Locality	binary	Urban; Rural	(Women living in urban vs. women who do not)
Marital status	categorical	Married; Living with partner, Separated/divorced, Widowed, Single	(Married women vs. non-married women)
Is disabled	binary	Yes; No	
Is employed	binary	Yes; No	
Has children	binary	Yes; No	
Generating an income	binary	Yes; No	
Living in HHs with loss in income	binary	Yes; No	

Annex 4

How to conduct the Hausman Test for Endogeneity:

The Hausman test for endogeneity is carried out in two steps20. Given the model researchers are trying to estimate:

$$y_i=\beta_0+\beta_1 x_i+Z\beta+u_i$$

Where:

y_i = main outcome of interest

and x_i = endogenous predictor of interest

and Z = a vector of exogenous predictors

First, regress x_i on Z and say instrument(s) are used (I) and save $\hat{v}_i$:

$$x_i=\pi_0+\pi_1 I_i+Z\pi+v_i$$

Because **Z** and **I** are exogenous (uncorrelated with u_i), it stands to reason that x_i is uncorrelated with u_i if and only if v_i is uncorrelated with u_i (the main predictor of interest x_i is exogenous if and only if the two error terms are uncorrelated). Therefore, it is essential to formally test whether these two error terms are correlated.

Second, run the Hausman test by regressing y_i on x_i, Z, and $\hat{v}_i$:

$$y_i=\beta_0+\beta_1 x_i+Z\beta+\delta\hat{\gamma}+\text{error}$$

For this test null hypothesis H_0: $\delta=0$. If the coefficient on $\hat{v}_i$ – that is, δ is significant, then it can be concluded that x_i was in fact endogenous because the two error terms were, in fact, correlated.

20 Wooldridge, J. 2003. Introductory Econometrics: A Modern Approach. 2nd ed. New York: Thomson Learning.

Annex 5

Regression results for countries

Dependent variables							
IPV before the COVID-19							
Counties	Albania	Bangladesh	Côte d'Ivoire	Jordan	Morocco	Nigeria	Paraguay
Treat	-0.020 (0.053)	0.135*** (0.048)	0.140*** (0.044)	0.001 (0.053)	0.330*** (0.057)	0.065 (0.041)	0.092 (0.058)
Constant	1.618*** (0.037)	1.160*** (0.034)	1.375*** (0.031)	1.199** (0.037)	1.481*** (0.040)	1.603*** (0.030)	1.656*** (0.041)
Observations	1,116	1,241	1,297	1,199	1,203	1,417	1,178
R2	0.0001	0.006	0.008	0.00000	0.027	0.002	0.002
Adjusted R2	-0.001	0.005	0.007	-0.001	0.027	0.001	0.001
Residual Std. Error	0.236 (df = 1114)	1.576 (df = 1239)	0.499 (df = 1295)	0.398 (df = 1197)	0.883 (df = 1201)	1.256 (df = 1415)	0.386 (df = 1176)
F Statistics	0.136 (df = 1; 1114)	7.820*** (df = 1; 1239)	9.920*** (df = 1; 1295)	0.0001 (df = 1; 1197)	33.815*** (df = 1; 1201)	2.468 (df = 1; 1415)	2.556 (df = 1; 1176)

Note: *p<0.1; **p<0.05; ***p<0.01

Dependent variables							
IPV after the COVID-19							
Counties	Albania	Bangladesh	Côte d'Ivoire	Jordan	Morocco	Nigeria	Paraguay
Treat	0.058 (0.052)	-0.022 (0.048)	0.232*** (0.049)	0.120** (0.054)	0.243*** (0.054)	0.163*** (0.049)	0.071 (0.057)
Constant	1.616*** (0.037)	1.423*** (0.034)	1.782*** (0.034)	1.169** (0.038)	1.620*** (0.038)	1.947*** (0.035)	1.745*** (0.041)
Observations	1,116	1,241	1,297	1,199	1,203	1,417	1,178
R2	0.001	0.0002	0.017	0.004	0.016	0.008	0.001
Adjusted R2	0.0002	-0.001	0.016	0.003	0.016	0.007	0.0005
Residual Std. Error	0.231 (df = 1114)	1.557 (df = 1239)	0.549 (df = 1295)	0.405 (df = 1197)	0.848 (df = 1201)	1.497 (df = 1415)	0.382 (df = 1176)
F Statistics	1.244 (df = 1; 1114)	0.208 (df = 1; 1239)	22.640*** (df = 1; 1295)	5.000** (df = 1; 1197)	19.932*** (df = 1; 1201)	11.073*** (df = 1; 1415)	1.565 (df = 1; 1176)

Note: *p<0.1; **p<0.05; ***p<0.01

220 East 42nd Street
New York, New York 10017, USA

data.unwomen.org
www.unwomen.org
www.facebook.com/unwomen
www.twitter.com/un_women
www.youtube.com/unwomen
www.flickr.com/unwomen
instagram.com/unwomen