CEPREMAP
CENTRE POUR LA RECHERCHE ECONOMIQUE ET SES APPLICATIONS

Document de travail (Docweb) n°2406

What one thinks, what one says and what one does: male justifications and practices of gender-based violence in Mali

Olivia BERTELLI, Thomas CALVO, Emmanuelle LAVALLÉE, Marion MERCIER, Sandrine MESPLÉ-SOMPS

Juin 2024

Ce qu'ils pensent, ce qu'ils disent et ce qu'ils font: justifications masculines et violences faites aux femmes au Mali [1]

Olivia BERTELLI[2], Thomas CALVO[3], Emmanuelle LAVALLÉE[4], Marion MERCIER[5], Sandrine MESPLÉ-SOMPS[6]

Résumé : Les violences basées sur le genre (VBG) sont largement répandues à travers le monde. Alors que la majorité de la littérature traite de la question en interrogeant les femmes en tant que victimes, cet article étudie la justification et la perpétration des VBG de la part des hommes. Cette étude porte sur le Mali l'un des pays avec les taux de prévalence de violences faites aux femmes les plus élevés au monde. Nous mesurons les taux de prévalence de huit opinions et comportements liés aux VBG auprès d'un échantillon représentatif de 1 200 hommes à Bamako. Nous administrons des questions directes standards ainsi qu'un protocole expérimental basé sur la méthode par comptage de réponses (aussi appelé en anglais « List Experiment » ou « Item count technique »). Cette méthode évite aux personnes interrogées de dévoiler directement l'information qui les concerne, et permet ainsi, en limitant toute forme d'intrusion, de mesurer les taux de prévalence exempts de biais déclaratifs.

Les taux de prévalence obtenus avec la méthode par comptage de réponses montrent qu'une large part de la population masculine justifie les VBG : presque un homme sur deux est en faveur de la mutilation génitale féminine (MGF) et justifie les violences conjugales ; un homme sur quatre déclare avoir déjà frappé physiquement une femme adulte. En comparant ces taux à ceux obtenus avec les questions directes, il apparait que plusieurs questions concernant la justification des VBG souffrent de biais de réponse significativement élevés lorsqu'elles sont posées de manière directe. En particulier, le soutien aux MGF est largement surestimé, montrant que le support pour ce type de pratiques discriminatoires est beaucoup moins répandu que ce que ce qui est déclaré dans les enquêtes directes standards. Au contraire, la justification des violences conjugales tend à être sous-estimée, possiblement en raison de la pression

[1]This work was made possible thanks to the funding of the French National Research Institute for Sustainable Development (IRD), Project GDRI–MOV (Groupement de recherche international - Sud « Mesurer et Observer les Violences, Violences de genre et de conflits dans les pays en voie de développement ») and the funding of the Centre for Economic Research and its Applications (CEPREMAP). We would like to express our gratitude to Massa Coulibaly and Moussa Coulibaly of the GREAT team for the survey management in Mali, and to Ousmane Z. Traore for the coding of the questionnaire. We sincerely thank all the enumerators. We are grateful to Hugo Colomb for excellent research assistance. We would also like to thank Katja Bergonzoli, Elise Huillery, Tista Mukherjee, as well as participants in the Nanterre workshop on Gender Issues and Development, PSL WiP workshop, Dauphine workshop on the Economics of Social Norms, in addition to participants at the CSAE conference and seminars at IE-UFRJ and ERUDITE for their very helpful comments and suggestions.

[2]DIAL-LEDa, UMR CNRS 8007 IRD 260, Université Paris-Dauphine, PSL Research University. Corresponding author. olivia.bertelli@dauphine.psl.eu

[3]DIAL-LEDa, UMR CNRS 8007 IRD 260, Université Paris-Dauphine, PSL Research University, IRD.

[4]DIAL-LEDa, UMR CNRS 8007 IRD 260, Université Paris-Dauphine, PSL Research University.

[5]DIAL-LEDa, UMR CNRS 8007 IRD 260, Université Paris-Dauphine, PSL Research University, CNRS.

[6]DIAL-LEDa, UMR CNRS 8007 IRD 260, Université Paris-Dauphine, PSL Research University, IRD.

sociétale croissante contre ce type de violences au Mali ces dernières années. Ces biais varient peu selon les caractéristiques des répondants (âge, éducation, groupe ethnique). Seule exception, les hommes ayant atteint un niveau d'éducation secondaire soutiennent nettement moins toutes les formes de VBG analysées dans cette étude que les hommes moins éduqués. En comparant nos résultats avec ceux d'autres contextes, il semble que le biais de réponse pourrait être influencé par le cadre juridique abordant les VBG et que les perceptions des individus sur les dimensions des violences « socialement acceptables » influencent leurs propres réponses aux questions directes, soulignant ainsi la nécessité de faire preuve de prudence quant à l'utilisation des données collectées via cette technique d'enquête.

Mots-clés : Violence basée sur le genre, attitudes, méthode par comptage de réponses, biais de réponse, Mali

What one thinks, what one says and what one does: male justifications and practices of gender-based violence in Mali

Abstract: Gender-based violence (GBV) is widespread across the world. While the majority of the literature focuses on women as the victims of GBV, this paper studies men's justifications for and perpetration of GBV in Mali, one of the countries with the highest GBV prevalence rates in the world. We elicit the prevalence rates of eight GBV-related opinions and behaviors among a representative sample of 1,200 men in Bamako, the capital city. We administer a list experiment and a set of direct questions to estimate response bias. The list experiment prevalence rates show that a large portion of the male population justifies GBV: nearly one in two supports female genital mutilation and intimate partner violence, and one in four has physically hit an adult woman. Moreover, several questions show significant response biases when asked with the standard direct question technique. Support for female genital mutilation is overestimated, indicating that it is less common than generally thought. Conversely, justification for intimate partner violence is underestimated, likely due to increased societal pressure against it in Mali. These biases vary little with respondent characteristics, although men with a secondary education level support all forms of GBV analyzed in this study the least. Comparing our results with those from other contexts suggests that response bias could be shaped by the legal framework addressing GBV and that people's perceptions of which dimensions are "socially acceptable" influence their own responses to standard direct questions, emphasizing the need to exercise caution with regard to the use of data collected via this survey technique.

Keywords : Gender-based violence, Attitudes, List Experiment, response bias, Mali

JEL Codes: J12, J16.

Introduction

Gender-based violence (GBV), in every shape and form, is both widespread globally and largely justified according to standard surveys. African countries report the highest prevalence of violence against women in terms of Intimate Partner Violence (IPV), child marriage and female genital mutilation (FGM) (UNSTAT, 2015, 2020). The use of violence against women is widely supported and justified by persistent discriminatory gender social norms rooted in patriarchal societies, even more so in low-income countries and especially in Africa (Jayachandran, 2015; Alesina et al., 2021).

The perpetration of GBV and its justification seem to be decreasing in some countries where data are available (UNSTAT, 2020). However, at a time when international organizations strongly campaign against these practices and governments are banning violence against women, people may be inclined to underreport their use and acceptance of GBV.[1] Social desirability bias and the fear of lack of confidentiality explain partly untruthful answers to sensitive questions (Tourangeau & Yan, 2007).

Moreover, recent work highlights the role of male social norms in changing attitudes and behaviors towards women (Jewkes et al., 2015; Bursztyn et al., 2020; Vaillant et al., 2020), pointing towards the need to better understand men's attitudes towards violence against women and prompting to include men when measuring and analyzing GBV practices and justification. In this study, we set out to examine male perspectives on GBV, one that is rarely explored. We assess male justifications for and use of GBV in terms of verbal, physical and sexual violence, as well as support for FGM and female early marriage among a representative sample of 1,200 adult males living in Bamako, the capital city of Mali. To account for possible declarative bias, we combine a List Experiment (LE) and a direct survey technique (DQ) to estimate response biases as well as arguably unbiased prevalence rates.

To our knowledge, this is the first analysis to provide accurate prevalence rates of men's GBV-related opinions and behaviors in West Africa. The context under study, Bamako, is the capital city of one of the countries with the worst conditions for women in the world, presenting among the highest rates of GBV victimization, FGM and child marriage. No law against domestic violence or FGM exists in Mali, and the legal age of marriage is 16 for women (vs. 18 for men).

Our main results show that men's justification of violence towards women is widespread, even netted out from response bias. Using the LE technique, about 44% of respondents report that a good wife is a circumcised wife, 54% are in favor of having their daughter excised, 43% justify IPV if a woman disrespects her husband, and one third thinks it is better for a girl to be married before 18 as well as normal for a man to control how his wife's money is spent. These results indicate a complex scenario regarding attitudes towards GBV. On the one hand, women's economic independence is relatively well accepted and female child marriage is not endorsed by the majority. On the other hand, support for FGM and IPV is widespread.

[1] For instance, research shows that in Ethiopia, where FGM has become illegal, true levels of support among both men and women for FGM is in fact much higher than self-reported in standard direct question surveys (De Cao & Lutz, 2018; Gibson et al., 2018).

Comparison of prevalence rates across the two survey techniques shows that three out of five self-reported opinions are significantly biased. Respondents underreport IPV justification (-10 percentage points) and massively overreport their support for wife and daughter excision (respectively, +24.5 and +20.7 percentage points). Interestingly, these response biases are fairly homogeneous in the study population, and do not vary according to respondent age, education level or ethnic group.[2] The fact that the true support for FGM is about 30% lower than openly declared is even more striking given that FGM prevalence rate in Mali is among the highest in the world. The significantly lower support we find for this discriminatory practice may signal room for policy reforms.

A second set of results concerns the self-reported perpetration of violence. We detect high prevalence rates, but find no significant response bias. About 26% of respondents declare having ever physically hit an adult woman, while nearly 2% report having ever raped a woman. These prevalence rates are close under the two survey techniques. A comparison of these results with those pertaining to the justification of GBV reveals that wife-beating is, on average, more justified (43%) than practiced (26%) by men, suggesting that justification of violence is not an ex-post rationalization of violent behaviors. Moreover, it could signal a change of social norms with regard to GBV and related practices in Malian society.

Our work contributes to the set of papers aiming to identify unbiased GBV prevalence rates - with the originality of studying potential perpetrators rather than victims. We add to the scarce body of work on misreporting GBV-related attitudes and behaviors by exploring men's attitudes and behaviors in a West African capital city. Adopting the men's perspective also aims at facilitating the design, implementation and acceptance of anti-GBV policies and encouraging the targeting of men and boys to prevent GBV. Contrasting our results with findings from other contexts suggests that response bias could be shaped by the legal framework addressing violence against women. In Mali, where the minimum legal age for girls to marry is 16, men do not underreport their support for child marriage, contrary to the findings of Asadullah et al. (2021) when questioning women in Bangladesh where the practice is illegal. In Ethiopia where FGM is banned, De Cao & Lutz (2018) and Gibson et al. (2018) find that support for FGM is underreported by men and women.[3] In contrast, we show that it is overreported by men in Mali, where FGM is legal. Moreover, with regard to GBV-related behaviors, the existing literature shows that GBV victimization is often misreported – mostly underreported – by victims (Phillips et al., 2010; Bulte & Lensink, 2019; Traunmüller et al., 2019; Lépine et al., 2020; Cullen, 2022). Among the few studies sampling men, Cullen (2022) finds that men in Rwanda underdeclare perpetrating emotional violence intended to limit their wife's contact with her family. Our results show instead that men in Bamako honestly answer direct questions about committing violence against women.

[2] The only notable difference pertains to men with low levels of formal education, whose IPV justification and support for men's control of their wife's monetary resources is more underreported using the standard DQ-technique than for the rest of the respondents.

[3] In Somalia, Gulesci et al. (2023) compare types of FGM reported by respondents with high and low social desirability index, and do not find any significant differential reporting bias in FGM status. They conclude that response bias may not be a significant concern in this country where the overall cutting rate is 98%, even though this practice is banned.

Our study also adds to an emerging literature evaluating interventions aimed at fighting violence against women by engaging boys and men (Hossain et al., 2014; Doyle et al., 2018; Vaillant et al., 2020). While most studies do not find a significant decrease in IPV prevalence, male self-reported attitudes towards IPV often appear to improve. Our study warrants about a possible social desirability bias reinforced by the intervention itself, possibly leading to an overestimation of the true intervention's impact on men's self-reported attitudes. Our results also call for caution in using self-reported data about GBV-related attitudes, especially for those studies examining the impact of laws banning excision (Camilotti, 2016; García-Hombrados & Salgado, 2023; Poyker, 2023) or punishing perpetrators of GBV (Sanin, 2021; Gulesci et al., 2024).

More broadly, this study contributes to the literature on the inaccuracy of standard direct question techniques for the measurement of sensitive attitudes and behaviors. The existing literature shows that misreporting is present in a wide range of sensitive issues: homophobia (Coffman et al., 2017), voting behavior (Holbrook & Krosnick, 2010; Rosenfeld et al., 2016), loan use (Karlan & Zinman, 2012), sexual behavior (Jamison et al., 2013; Coffman et al., 2017; Plummer et al., 2004), the use of condoms among female sex workers (Treibich & Lépine, 2019; Lépine et al., 2020) and child labor (Jouvin, 2023).[4]

Furthermore, we add to the recent debate about the capacity of LEs in measuring unbiased prevalence estimates and their relevance when response bias is small. Lépine et al. (2020) discuss the trade-off between potential bias reduction from minimizing misreporting, and efficiency loss given LE's higher variance. Blair et al. (2020) highlight that bias in sensitive items in political science are often small – around 5 percentage points – limiting the contribution of LEs. We take part in this debate by showing that, when it comes to Malian men's support for GBV, response bias can in fact be substantial - consistent with the results of Cullen (2022), Lépine et al. (2020) and Asadullah et al. (2021) in different frames.

Finally, this paper contributes to the debate about the heterogeneity of response bias.[5] From an individual perspective, we find that misreporting varies very little across respondent characteristics, with the exception of support for IPV and for husband's control that are more underreported by men with lower levels of formal education.

The rest of the paper is structured as follows. Section 1 presents the sample selection, the randomization design and the lists of items under scrutiny; and provides evidence supporting the validity of the experimental design. In Section 2 we outline the empirical specifications for data analyses. Section 3 reports the results and Section 4 provides several robustness checks. Section 5 draws the main conclusions.

[4] For a recent review of LEs in political science, see Blair et al. (2020).

[5] In a South Central Ethiopian community, Gibson et al. (2018) find that support for FGM is mostly underreported by educated respondents while support to IPV is mostly underreported by uneducated adults. In the Afar region in Ethiopia, De Cao & Lutz (2018) observe that uneducated women are particularly prone to underreporting support for FGM. In contrast, women's education is positively related to the underreporting of IPV in Kerala (Joseph et al., 2017), among female micro-credit clients of Lima's poor districts in Peru (Agüero & Frisancho, 2022) and in the Kebby state of Nigeria (Cullen, 2022). Lastly, Joseph et al. (2017) also observe that the youngest and oldest respondents underreport domestic violence more than those of the middle-age cohort.

1 Context and survey experiment

This study was carried out in Bamako, the capital city of Mali, one of the lowest ranked countries in world in terms of the Human Development Index (188th out of 193 countries) and the Gender Inequality Index (172th out of 181 countries) according to the last UNDP Human Development Report (UNDP, 2024). According to the latest available Demographic and Health Survey data (2018), almost one Malian woman out of two has been victim of GBV in her lifetime, FGM rates spike at 90% and female child marriage (i.e., before the age of 18) is widespread (54%). Discriminatory gender norms are also widely accepted as 79% of women justify wife-beating, the second highest rate in the world after Afghanistan (UNSTAT, 2020).

While no laws punishing FGM and IPV exist today in Mali and the minimum legal age for girls to marry is 16, in the past few years the civil society has risen multiple times claiming policy reforms for punishing the perpetrators of violence against women. Moreover, over the years, numerous efforts have been made to implement a ban on FGM in Mali. Various draft laws prepared by both government departments and non-governmental organizations have failed to gain approval and pass through parliament (28toomany, 2018).

1.1 Sampling design and randomization

The face-to-face survey was carried out in November 2022 by our local research partner GREAT (*Groupe de recherche en économie appliquée et théorique*) among 1,200 adult men representative of the adult male population in Bamako. The survey was conducted by a team of local male enumerators and took place in 150 enumeration areas (EA) of the six *communes* of the capital city. The EAs were randomly selected from the list established by the National Statistics Office (INSTAT) during the mapping phase of the last General Population Census in 2022. Within each EA, a random walk protocol was implemented to select the sampling households. In each household, enumerators randomly selected one permanent adult male member speaking Bambara or French to be surveyed.[6]

Respondents were randomly assigned to the treatment or control group in the following way. Each enumerator was assigned to interview 100 men, half of whom were assigned to the treated group and half to the control group. All enumerators alternated control and treatment questionnaires. The first questionnaire was a control one if the enumerator ID number was even, while a treatment questionnaire was used if the enumerator ID number was odd.[7] The difference between the two groups relies only on the types of lists administrated. The treatment group answers the sensitive items using eight lists. The control group has the same eight lists but without any sensitive item, that are instead administrated with a standard direct-question technique.

[6] The selected individuals were given the opportunity to reschedule the interview to another date. Those refusing the interview were replaced by another man living in a different household selected in the same EA using the same random walk protocol. Bambara and, to a lesser extent, French are the vernacular languages in Mali.

[7] This allocation of respondents per enumerator avoids potential biases between the treatment and the control groups resulting from differences in the way the interviewers administered the questionnaires.

Table 1: Randomization balance checks.

	Control			Treatment			
	n	mean	sd	n	mean	sd	Diff
Age	600	37.76	15.99	600	36.76	15.41	-0.994
Primary education achieved	600	0.98	0.13	600	0.98	0.12	0.002
Secondary education achieved	600	0.67	0.47	600	0.69	0.46	0.021
Married	600	0.61	0.49	600	0.59	0.49	-0.024
Never married	600	0.32	0.47	600	0.35	0.48	0.035
Spouse lives in HH	366	0.92	0.28	353	0.92	0.27	0.003
At least one adult woman in HH	600	0.56	0.50	600	0.54	0.50	-0.019
High FGM prevalence ethnic group	600	0.77	0.42	600	0.76	0.43	-0.015
Mother tongue Bambara	600	0.66	0.47	600	0.67	0.47	0.003
HH head	600	0.47	0.50	600	0.47	0.50	0.003
HH head's father	600	0.03	0.18	600	0.03	0.18	-0.001
HH head's son	600	0.25	0.43	600	0.27	0.45	0.024
HH head's relative	600	0.23	0.42	600	0.21	0.41	-0.021
HH member unrelated	600	0.02	0.15	600	0.02	0.13	-0.005
Joint F-test					0.56		
p-value					0.848		

Notes: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Results reported in Table 1 show that respondents assigned to the treatment and control groups are not statistically significantly different across a set of observable characteristics. It is worth noticing that almost all respondents completed at least primary education while about two thirds completed secondary education, 60% are married and more than half live with an adult woman in the household. High FGM prevalence ethnic groups correspond to groups (based on mother tongue) with rates of genital cutting close to 90% as documented in Diabate & Mesple-Somps (2019). They include (among others) the Bambara, Malinke, Fulani and Senufo.

1.2 Lists of items

In a list experiment, respondents are randomly allocated to a control and a treatment group. They are given a list of statements and asked to declare how many of these statements they experienced (or agree with), without specifying which ones. In the control group, respondents are questioned based on a list of J so-called "baseline" items, while in the treatment group respondents are questioned referring to a list of $J + 1$ items, the additional item being the sensitive one representing the research object. Given the random allocation of respondents into the two groups, the prevalence rates of the baseline items should be equal across groups. The difference between the average number of items declared in the treatment group and in the control group thus provides an estimate of the prevalence of the sensitive item in the sample.

Our survey contains eight list experiments, each one with a different sensitive item about a specific GBV-related behavior or opinion. The eight sensitive items were also administrated to the control group under the form of a direct "Yes/No" question. The sensitive questions are

phrased in such a way that the "Yes" answer indicates having committed or justifying the act of violence to which the sensitive item refers to. The sensitive items, as formulated in the lists or directly asked about, are reported in Table 2.[8] Their rank in the list was randomly determined. Three items concern the perpetration of violence (verbal, physical, and sexual violence), while the other five are related to opinions in favor of GBV (economic control, female child marriage, female genital cutting, and physical intimate partner violence).

The survey design thus allows us to compute two measures of the prevalence rate of each sensitive item: one from the LE, based on the comparison between the counts declared by respondents in the treatment and control groups; and one from the direct questions administrated to the control group.

Table 2: Items and questions about gender-based violence behaviors and opinions.

	Item (LE)	**Direct question (DQ)**
I1. Verbal violence	Sometimes my words make my wife cry	Do your words ever make your wife cry?
I2. Physical violence	I have already hit one or more women	Have you ever hit one or more women?
I3. Rape	I have already forced a woman to have sex with me	Have you ever forced a woman to have sex with you (although she didn't want to)?
I4. Husband's control	It's up to the man to decide how his wife's money should be spent	Is it up to the man to decide how his wife's money should be spent?
I5. Child marriage	It's better for a girl to get married before the age of 18	Is it better for a girl to get married before the age of 18?
I6. FGM wife	A good wife is an excised wife	Is a good wife an excised wife?
I7. FGM daughter	If I were to have a daughter today, I would have her circumcised	If you had a daughter today, would you have her circumcised?
I8. IPV justified	A wife who does not respect her husband deserves to be hit	Does a wife who doesn't respect her husband deserve to be hit?

1.3 Validity of the list experiment

The validity of a list experiment rests on the no-liar assumption, which implies the absence of any floor or ceiling effects. This implies in particular the assumptions of no floor or ceiling effects. The LE also relies on the hypothesis of no design effect. We follow Blair & Imai (2012) and Glynn (2013) to explore the plausibility of these assumptions in our data.

A floor effect is likely to occur when respondents disagree with all the baseline items. Two scenarios would then bias treated respondents' answers in the list experiment: (i) if they agree with the sensitive item only, but fear their true preference will be revealed by reporting 1, they could instead choose to report 0. This would yield an underestimation of the LE prevalence rate; (ii) on the contrary, if they disagree with all items including the sensitive one, they could be reluctant to answer 0 – which mechanically reveals their opinion – but rather prefer to hide their true preference by answering 1, yielding an overestimation of the LE prevalence rate.

[8] Table A1 in the Appendix also presents the non-sensitive items.

A ceiling effect, by contrast, takes place when respondents agree with all items including the sensitive one. While answering $J + 1$ (with J being the number of baseline items in the list) would reveal their response to each item, respondents may prefer to answer J to hide their true preference for the sensitive item, leading to an underestimation of the LE prevalence rate.

To minimize the risk of ceiling and floor effects, LEs should avoid baseline items either too rare (that would result in a large share of "No" answers) or too common (that would result in a large share of "Yes" answers). In order to identify baseline items limiting floor and ceiling effects, we conducted a pilot survey in October 2022.[9] A number of possible baseline items were administered under the form of direct "Yes/No" questions. In this way, we identified high-prevalence and low-prevalence items, as well as possibly sensitive items (generating missing answers when asked directly), allowing us to build the sets of baseline items to be combined with each sensitive item in the list experiment. More specifically, we excluded items that appeared to be sensitive or that were not generating enough variance, and then designed the lists accounting for correlations between baseline items following the guidelines of Chuang et al. (2021). Appendix Table B1 shows descriptive statistics of the selected baseline items asked directly during the pilot survey, whereas Table B2 reproduces coefficients of correlation between them.[10] The pairwise correlations between items are not too large and, for most of the lists, are both positive and negative, which is recommended in order to limit the risks of floor and ceiling effects.

We explore whether we face such risks in our final data by looking at the shares of individuals agreeing with j items for each list (with $j = 0, 1, 2, 3$ for the controls and $j = 0, 1, 2, 3, 4$ for the treated). Figure 1 is a graphic representation of these shares by treatment status (see Table C1 in Appendix C for the precise proportions). A floor effect might occur when respondents disagree with all the baseline items, resulting in a high prevalence of 0 among the control respondents. On the other hand, a ceiling effect may occur when respondents agree with all the baseline items, resulting in a high prevalence of 3 in the control group. Attention should thus be paid to the tails of the distributions in the control group.

Reassuringly, the share of respondents in the control group that do not conform with any baseline item is always lower than the share of respondents agreeing with one or two items. It is below 5% in the majority of lists (lists 1, 2, 4, 5 and 8) and below 15% in two other lists (lists 3 and 6), in line with results from other studies. It is, however, higher in list 7, where 20% of the control respondents do not agree with any item, possibly signaling a floor effect. As explained above, in such a case, treated respondents who would like to answer 1 may declare 0 if they fear that answering 1 would signal that they agree with the sensitive item, whereas those who would like to answer 0 may prefer to answer 1 to conceal their true preference. The share of treated respondents answering 0 is equal to 8% in list 7, which suggests that if the former effect is at play, it is of relatively small magnitude. However, we cannot fully exclude that some treated respondents answered 1 instead of 0 to hide their disagreement

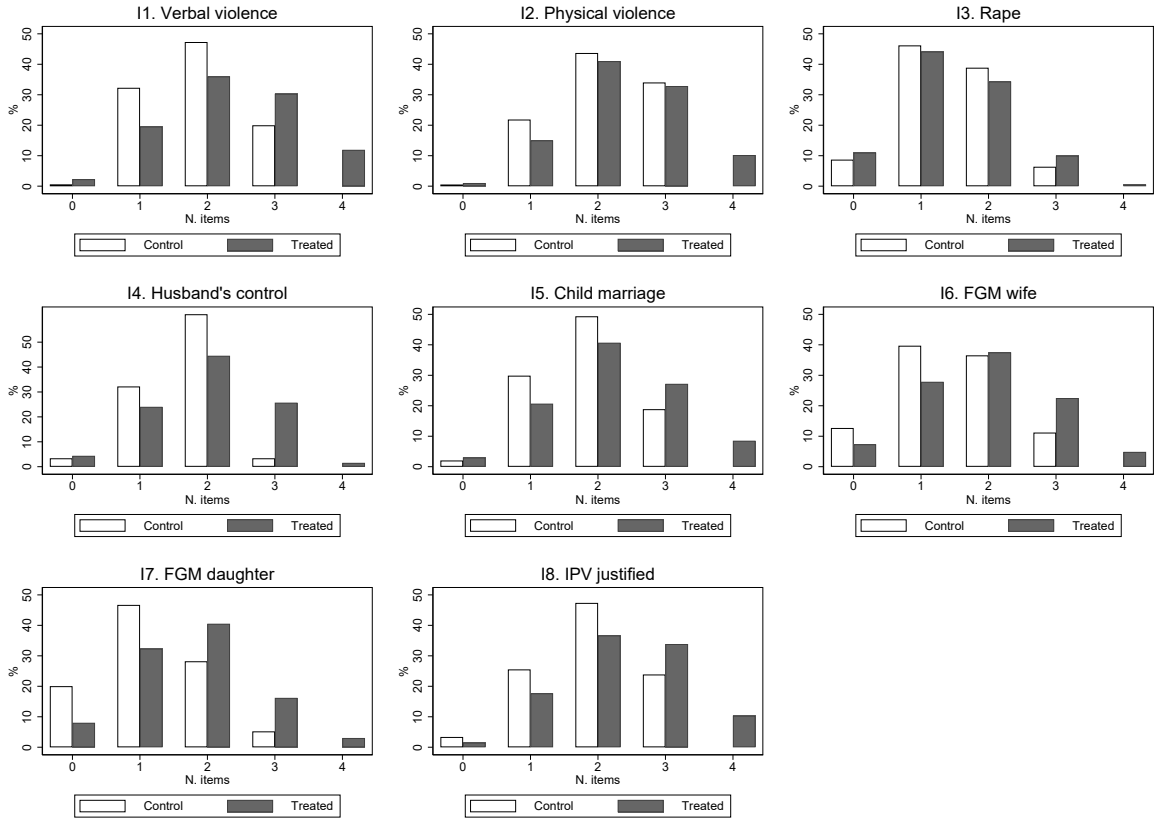
[9] See Appendix B for a detailed presentation of the pilot.

[10] Note that five baseline items present in the final lists were not tested under the form of direct questions in the pilot, because we replaced some of the items from the pilot that were unsatisfactory with new ones, and because list number 4 (attitudes to women's financial autonomy) was added to the survey after the pilot was completed.

with the sensitive item, thus yielding an upward bias in the prevalence rate of daughter genital mutilation acceptance measured with the LE – a risk that will be kept in mind when interpreting our results. Nevertheless, the correlations between answers to the baseline items from the pilot survey, which are not too high and both positive and negative, somehow mitigate the concern of a floor effect affecting list 7 (see Table B2 in the Appendix).

Turning to the risk of ceiling effect, lists 1, 5 and 8 present shares of respondents in the control group agreeing with all three baseline items above usual thresholds (between 18% and 24%). However they remain below the share of respondents who agree with one or two items. These shares are even smaller in lists 3, 4, 6 and 7 (between 3 and 11% of respondents in the control group). The share of respondents agreeing with all three baseline items in list 2 is greater than the share of those agreeing with one item only, but lower than the share of those agreeing with two items. It suggests a possible ceiling effect, with treated respondents answering 3 rather than 4 to hide their agreement with the sensitive item, which would lead us to underestimate the prevalence rate of physical violence.

Figure 1: Average number of items and share of respondents agreeing with j items, by list and treatment status.



The validity of a list experiment also rests on the assumption that there are no design effects. These arise when the inclusion of a sensitive item affects the answers to the baseline items. To explore the risk of design effect, we compare the proportions of respondents in the treatment and control groups reporting *at least* J ($J = 1, 2, 3$) items. We expect a higher share of respondents

declaring at least one/two/three items in the treatment group, given the longer list of items compared to the control group. A negative difference between the treated and control shares would signal the presence of a design effect (Glynn, 2013). Though Table 3 shows a small negative difference in the proportions of respondents reporting at least one item for lists 1 to 5, which could yield a downward bias in the estimated LE prevalence, none of these differences are significant, except for list 1, significant at 10% level.

We run two more tests to examine whether these negative differences are due to luck or to the list experiment design. The first one tests the null hypothesis that none of the joint probabilities Pr(R=r,S=0) (where R are the baseline items and S is the sensitive item) are smaller than zero. The second one tests the null hypothesis that none of the joint probabilities Pr(R=r,S=1) are smaller than zero. The rejection of one or the other of these two hypotheses would signal a risk of design effect. We report standard p-values and Bonferroni-adjusted p-values for each test and each item in Appendix Table C2. The results show that we cannot reject the hypothesis of no-design effect for all items, except item 1. Based on these results and those of Table 3, we prefer to exclude item 1 from the rest of the analysis.[11]

Table 3: Share answers by item and treatment status.

	I1	I2	I3	I4	I5	I6	I7	I8
At least one item (T)	98.67	99.00	89.00	95.67	97.00	92.67	92.00	98.50
At least one item (C)	99.67	99.50	91.33	96.67	98.00	87.33	80.00	96.67
Diff. T-C	-1.00*	-0.50	-2.33	-1.00	-1.00	5.33	12.00	1.83
At least two items (T)	87.17	84.00	44.83	71.67	76.33	64.83	59.67	80.83
At least two items (C)	80.00	77.67	45.17	64.50	68.17	47.67	33.33	71.17
Diff. T-C	7.17	6.33	-0.33	7.17	8.17	17.17	26.33	9.67
At least three items (T)	66.00	43.00	10.50	27.17	35.67	27.33	19.17	44.17
At least three items (C)	51.17	34.00	6.33	3.33	18.83	11.17	5.17	23.83
Diff. T-C	14.83	9.00	4.17	23.83	16.83	16.17	14.00	20.33

*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

2 Empirical strategy

The first empirical step of our analysis consists in obtaining the prevalence rates measured under the two survey techniques for each sensitive item and in comparing them to estimate possible response bias.

The prevalence rate of sensitive item k measured with the DQ technique, $\hat{g}_k$, can be obtained with a regression of the direct answers on a constant:

$$Z_{k,i} = g_k, \tag{1}$$

where $Z_{k,i}$ equals 1 if individual i answers "Yes" to sensitive question k, and 0 otherwise.

[11] We also test the hypothesis of no-design effect for each sub-group under study in the heterogeneity analysis. We cannot reject the hypothesis of no-design effect for all items and sub-groups, except for men with secondary education and under 25 years old in item 3. Results are available upon request.

The prevalence rate of sensitive item k using the LE technique, $\hat{\gamma}_k$, is measured based on the comparison of the control and treatment groups. In particular, we estimate the following equation:

$$Y_{i,k} = \alpha_k + \gamma_k T_i + \lambda_k W_i^{'} + \epsilon_{i,k} \tag{2}$$

where $Y_{k,i}$ denotes the response given by individual i to the list of items containing k ($Y_{k,i}$ ranges from 0 to 3 for respondents of the control group and from 0 to 4 for respondents of the treatment group), T_i is a dummy variable equal to 1 if i is in the treatment group and to 0 if i is in the control group, and $\epsilon_{i,k}$ is the error term. We introduce a vector of control variables $W_i^{'}$, that includes enumerator fixed effects to account for potential systematic differences across enumerators, along with a number of respondent characteristics – namely, a set of binary variables: having completed secondary education, being from a high FGM-prevalence ethnic group, a set of age dummies as well as of dummies characterizing the respondent's relationship with the household head. Equation 2 yields an accurate estimate of the prevalence of item k provided that the assignment to treatment is random.

To detect response bias, we test whether $\hat{g}_k$ and $\hat{\gamma}_k$ are significantly different from each other. For each sensitive item, the difference between the LE and DQ prevalence rates can be interpreted as the size of the declarative bias the LE-technique neutralizes, showing the value-added of using this technique rather than the DQ.

The second empirical step of our analysis consists in exploring whether response biases vary across respondent characteristics – namely, education, age, and ethnic group. Sociodemographic characteristics are likely to influence the social norms to which individuals aim to comply, the degree to which they are attached to these norms, the feeling of shame that they associate with certain behaviors or how sensitive they are to societal pressure against GBV – hence the social desirability bias. Individual characteristics might also determine their fear of a lack of confidentiality, their perception of the possible consequences of their answers being disclosed to third parties, and/or their willingness to keep their personal information private.

Following Blair & Imai (2012), we compute prevalence rates by education level, age category (below or above 25 years old), and ethnicity. We differentiate between groups with high and low FGM prevalence rates based on Diabate & Mesple-Somps (2019). Based on data from 2009, they observe that the rate of FGM among girls below the age of 14 is close to or even surpasses 90% in the Bambara, Soninke, Malinke, Senufo, Fulani and Dogon groups, while it is less than 50% among Bobo and Songhai girls.

The DQ prevalence rate of each sensitive question is estimated using the following equation:

$$Z_{i,k} = g_k + \mu_k X_i + \epsilon_{i,k} \tag{3}$$

where $Z_{i,k}$ denotes the answer given by individual i to sensitive question k and X_i is one of the dimensions under scrutiny captured with a dummy variable (secondary education, below 25 years of age, high FGM-prevalence ethnic group). The DQ prevalence rate of sensitive question k is then given by $\hat{g}_k$ for the omitted category of variable X_i, and by $\hat{g}_k + \hat{\mu}_k$ for the category under study.

The LE prevalence rate of each sensitive question is estimated with the following equation:

$$Y_{i,k} = \alpha_k + \beta_k X_i + \gamma_k T_i + \delta_k X_i \times T_i + \lambda_k W_i^{'} + \epsilon_{i,k} \tag{4}$$

where we interact the treatment status T_i with one of the dummies of interest. The estimated term $\hat{\gamma_k} + \hat{\delta_k}$ captures the LE prevalence rate among individuals with $X_i = 1$ while $\hat{\gamma_k}$ measures the prevalence for those with $X_i = 0$. As before, $W_i^{'}$ is the vector of enumerator fixed effects and individual controls.

To assess whether response bias varies across respondent characteristics, we compare $\hat{g}_k$ and $\hat{\gamma_k}$ on the one hand, and $\hat{g}_k + \hat{\mu_k}$ and $\hat{\gamma_k} + \hat{\delta_k}$ on the other, based on the coefficients estimated in Equations 3 and 4.

In all our estimations, we cluster standard errors at the enumeration area-level.

3 Results

In this section we first present the results on GBV justification, and then examine whether misreporting GBV justification is homogeneous in our sample. Lastly, we turn to the results on GBV perpetration.

3.1 The justification of GBV

Table 4 presents the average prevalence rate of each sensitive item related to the justification of GBV, measured with the two survey techniques (*i.e.*, $\hat{g}_k$ estimated from Equation 1 and $\hat{\gamma_k}$ estimated from Equation 2), along with the difference between these two. The information declared under DQ suffers from a reporting bias if the difference between LE and DQ is significantly different from zero.

Focusing first on the prevalence rates revealed by the LE, we observe that a large share of respondents justify GBV practices. About 54% of men would have their daughter genitalia cut if they had one, about 44% consider that a good wife is an excised wife, and 43% justify IPV if a wife disrespects her husband. The dimensions that encounter the least support are female child marriage (31%) and husband's control of his wife's monetary resources (32%). These results show a relatively good acceptance of women's economic independence and of increasing girls' age of marriage at 18 years old. In contrast, support for FGM and IPV appears to be still widespread.

Some of these prevalence rates are subject to a declarative bias when they are measured with direct questions. In particular, we find substantial biases for items relating to the support for FGM and for IPV justification.

When asked directly, more than two-thirds of respondents approve of FGM: 68% declare that a good wife is an excised wife and 75% report that, if they had a girl, they would have her cut. These prevalence rates are remarkably close to those found in the latest DHS data for

Mali (2018) where 72% of men in Bamako declare that female excision should continue. For both questions, the bias is large and statistically significant: support for FGM is overreported by 25 and 21 percentage points, respectively, corresponding to 37% and 28% of the respective DQ-prevalence rates.[12]

This considerable misreport provides a sense of how difficult it is to speak about FGM for the male population of Bamako, and suggests that openly expressing reservations about FGM practices is problematic. These results differ from those of De Cao & Lutz (2018) and Gibson et al. (2018), who find that support for FGM is underreported when elicited via standard direct questions in rural communities in Ethiopia, where FGM is illegal and gradually disappearing. By contrast, in Mali where FGM is not illegal and still widely practiced (though with a varying intensity across ethnic groups), support for FGM is overreported in DQ. A possible explanation for individuals being in fact much less in favor of FGM than they actually declare, is the (mis)perception of the prevalent FGM social norm, resulting in a substantial social desirability bias and pointing towards a possible pluralistic ignorance with regard to the true social norm.

With regard to IPV justification, our results also emphasize a significant response bias. When asked directly, 33% of men justify IPV when a woman disrespects her husband.[13] This proportion reaches 43% in the LE – the 10 percentage point bias being statistically significant and corresponding to 30% of the DQ prevalence rate. Answers obtained under DQ suffer in this case from a negative bias, suggesting that, contrary to support for FGM, it is not socially desirable for Malian men to openly justify IPV. This result is in line with the study by Gibson et al. (2022) who find a considerable underreporting of IPV justification among men in Ethiopia – though statistically significant only for men without secondary education.

It therefore appears that opinions on both FGM and IPV are sensitive subjects, leading to biases in responses to directly asked questions. This leads us to exercise caution with regard to using data collected using a standard direct technique, and perhaps calls into question the robustness of the studies examining the impact of laws banning excision (Camilotti, 2016; García-Hombrados & Salgado, 2023; Poyker, 2023) or punishing perpetrators of GBV (Gulesci et al., 2024; Sanin, 2021).

Finally, neither the support for husband's control nor for early marriage appear to be affected by response biases. This suggests that, while prevalence rates are relatively high (around 31%), revealing information about these personal opinions is not perceived as sensitive.[14] As far as support for child marriage is concerned, the absence of response bias differs from the results of Asadullah et al. (2021) who document underreporting in Bangladesh – a country where child marriage is banned, contrary to Mali where girls are legally authorized to get married from the age of 16.

[12] As discussed above, a possible floor effect concerning the item on daughter FGM could lead to an overestimation of the LE-based prevalence rate. The response bias would then be underestimated.

[13] Again, the measured prevalence rate under DQ is close to the one measured in the latest available DHS data (2018): 26% of men in Bamako (31% of men in urban areas) justify wife beating – a prevalence rate which is consistent with what we observe under DQ, given that the question is phrased differently.

[14] Unfortunately, we cannot compare these prevalence rates with those found by other sources, as these questions are not collected in DHS-type surveys.

As shown in Table 4, the results are robust to p-value correction for false discovery rate (Benjamini et al., 2006; Anderson, 2008). Besides, it is worth mentioning the trade-off between bias reduction and precision loss that characterizes LEs. In fact, while they are meant to produce unbiased estimates, they also increase the variance of estimates. Following Blair et al. (2020) and Lépine et al. (2020), we estimate the minimum sample size for which our list experiments are likely to produce more valid results than direct questions. Results reported in Appendix D, Table D1 show that our sample size, given the two survey techniques prevalence rates and baseline items' variance, is adequate for detecting biases for all items with regard to justification of GBV except for husband's control over his wife's monetary resources.

Table 4: GBV justification: prevalence rates and response biases

	(1) Husband's control	(2) Child marriage	(3) FGM wife	(4) FGM daughter	(5) IPV justified
LE	0.317***	0.313***	0.435***	0.538***	0.434***
	(0.042)	(0.055)	(0.054)	(0.050)	(0.050)
DQ	0.325***	0.382***	0.680***	0.745***	0.333***
	(0.020)	(0.020)	(0.018)	(0.019)	(0.019)
LE-DQ	-0.008	-0.069	-0.245	-0.207	0.100
p-value	0.852	0.209	0.000	0.000	0.048
Sharpened q-values	0.515	0.270	0.001	0.001	0.106
N LE	1200	1200	1200	1200	1200
N DQ	597	591	556	569	594

Notes: OLS estimations; LE equations include enumerator and age fixed effects, as well as dummies denoting secondary education level, belonging to an ethnic group characterized by a high prevalence of FGM, and the relationship with the household head.
Robust standard errors in parentheses (adjusted for clustering at enumeration area level); *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

3.2 Heterogeneity in GBV justification and its misreporting

We investigate whether responses and response biases regarding the justification of GBV vary according to respondents' education, age and ethnicity. Specifically, we distinguish men having achieved at least secondary level of education and those with a lower level of education; men aged under or over 25; and men belonging to a group where FGM prevalence is high or low.[15] Columns 1 and 2 of Table 5 show the LE-based prevalence rates of the omitted and non-omitted categories, and Column 3 the difference between them. Columns 4 and 5 display the reporting biases, namely the difference between the LE-based and the DQ-based prevalence rates for the omitted category ($\hat{\gamma}_k - \hat{g}_k$) and the non-omitted category ($(\hat{\gamma}_k + \hat{\delta}_k) - (\hat{g}_k + \hat{\mu}_k)$), respectively. As explained in Section 2, the analysis relies on the comparison between the results of the estimation of Equations 3 and 4. The significance levels of the differences based on the p-values and the adjusted q-values are reported, respectively, in superscript and in squared parentheses.

Looking at the prevalence rates measured with the list experiments reveals a number of

[15] Results do not vary with respect to whether the respondent is the household head, he completed primary school, he is married and whether the married respondents are in a polygamous or monogamous marriage. We do not have the information on whether respondents have at least one daughter.

interesting patterns (columns 1-3 of Table 5). As highlighted above, the form of GBV that is still most widely approved is FGM, and this is true across all respondent characteristics.

The widest prevalence gap for GBV justification is observed between men educated to secondary level and those with a lower level of education, whereas young and older men, or those from high- and low-FGM prevalence ethnic groups, exhibit similar levels. In particular, men having achieved a secondary level of education are much less likely to be in favor of a husband's control of their wife's money (25.7% against 44.1% among men without secondary education). They are also less likely to prefer a genitally mutilated wife (36.5% against 58.1%), support the genital mutilation of daughters (48% against 65.9%) or justify IPV (34.9% against 60.8%). Support for child marriage is also lower among men educated to secondary level, though the difference is not statistically significant. It is also worth mentioning that, strikingly, the younger generation does not appear to approve less GBV than the older generation. Prevalence rates among men under or above 26 years old are very close to one another and younger men are even slightly more supportive of discriminatory norms than men with secondary education (column 2 of Table 5).

Misreporting appears homogeneously distributed in our sample. Irrespective of respondent education, age or ethnic group, support for FGM is consistently overreported when directly asked. A slight nuance may be noted concerning support for wife and daughter FGM, with an overreporting bias that becomes insignificant for men without secondary education when correcting p-values for false discovery rate. Misreporting of wife FGM is thus weak among respondents with lower levels of education. On the other hand, the underreporting of IPV justification is mostly driven by men without secondary education.

Table 5: Prevalence rates and response biases across sub-groups

		Prevalence rate			Response bias (LE-DQ)	
		0	1	Difference	0	1
Item	Characteristic	(1)	(2)	(3)	(4)	(5)
Husband's control	Secondary education	0.441	0.256	-0.185**	0.133	-0.077
		(0.077)	(0.047)	(0.091)	[0.248]	[0.279]
	Under 26 years old	0.342	0.257	-0.084	0.016	-0.066
		(0.047)	(0.077)	(0.089)	[0.731]	[0.537]
	High FGM prevalence ethnic groups	0.363	0.302	-0.061	0.025	-0.019
		(0.084)	(0.049)	(0.100)	[0.731]	[0.731]
Child marriage	Secondary education	0.345	0.298	-0.047	-0.050	-0.078
		(0.088)	(0.063)	(0.105)	[0.607]	[0.375]
	Under 26 years old	0.309	0.323	0.014	-0.115*	0.049
		(0.058)	(0.095)	(0.105)	[0.180]	[0.644]
	High FGM prevalence ethnic groups	0.445	0.272	-0.173	0.026	-0.099
		(0.109)	(0.056)	(0.118)	[0.757]	[0.248]
FGM wife	Secondary education	0.581	0.365	-0.216*	-0.157*	-0.284***
		(0.089)	(0.065)	(0.112)	[0.244]	[0.001]
	Under 26 years old	0.478	0.334	-0.144	-0.215***	-0.313***
		(0.064)	(0.096)	(0.118)	[0.010]	[0.010]
	High FGM prevalence ethnic groups	0.402	0.446	0.044	-0.279**	-0.234***
		(0.112)	(0.055)	(0.119)	[0.074]	[0.002]
FGM daughter	Secondary education	0.659	0.480	-0.179*	-0.173*	-0.222***
		(0.085)	(0.058)	(0.102)	[0.178]	[0.003]
	Under 26 years old	0.559	0.490	-0.068	-0.205***	-0.209***
		(0.061)	(0.073)	(0.093)	[0.010]	[0.023]
	High FGM prevalence ethnic groups	0.456	0.564	0.108	-0.245**	-0.194***
		(0.105)	(0.054)	(0.116)	[0.107]	[0.006]
IPV justified	Secondary education	0.608	0.349	-0.259***	0.265***	0.021
		(0.078)	(0.060)	(0.097)	[0.010]	[0.731]
	Under 26 years old	0.422	0.460	0.037	0.091	0.120
		(0.056)	(0.089)	(0.103)	[0.261]	[0.309]
	High FGM prevalence ethnic groups	0.400	0.444	0.044	0.091	0.103*
		(0.084)	(0.057)	(0.099)	[0.459]	[0.192]

Notes: Column 1 reports the LE-based prevalence rate of the subgroup of the omitted category ($\hat{\gamma}_k$), while column 2 reports that of the non-omitted subgroup ($\hat{\gamma}_k + \hat{\delta}_k$). Column 3 reports the difference between the prevalence rates for the omitted and non-omitted category ($\hat{\delta}_k$). Column 4 reports the difference between the LE-based and the DQ-based prevalence rate for the subgroup of the omitted category, while column 5 reports the same difference for the non-omitted subgroup.

Secondary education refers to the achievement of a secondary education level; Under 25 years old equals one if the respondent is under 25 years old; High FGM prevalence ethnicity equals one if the respondent belongs to a high FGM prevalence ethnic group.

OLS estimations; LE equations include enumerator and age fixed effects, as well as dummies denoting secondary education level, belonging to an ethnic group characterized by a high prevalence of FGM, and the relationship with the household head.

Robust standard errors clustered at the EA level in parentheses. Adjusted q-values correcting for false discovery rate following Benjamini et al. (2006) and Anderson (2008) are in squared brackets.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

3.3 GBV practices

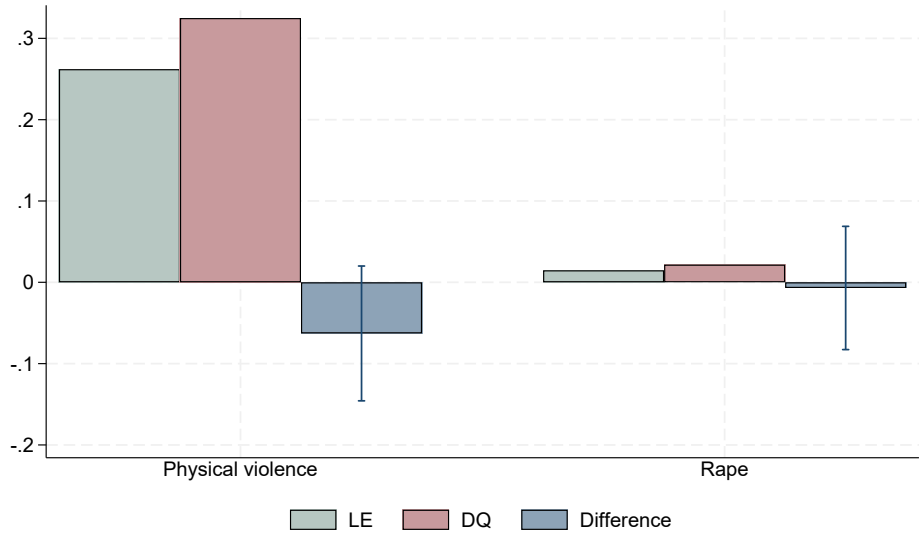
We further investigate whether the reporting of actual behavior, i.e., the perpetration of violent acts against women, suffers from response bias. Even though it was hazardous to phrase the behavioral-related sensitive direct questions and items as symmetric matches of the attitude-

related ones, we can nevertheless account for two violent behaviors: hitting and raping women.[16]

Somewhat surprisingly, self-reported perpetration of physical violence or rape against women does not appear to be affected by misreporting (see Figure 2 and Table E1 for detailed results). The LE and DQ techniques yield similar prevalence rates: about 26% of respondents have ever committed physical violence against an adult woman, and nearly 2% have ever raped a woman. It is important to mention that, as discussed in Section 1.3, a possible ceiling effect may lead to an underestimation of the LE prevalence of physical violence, further reducing the size of the estimated response bias.[17]

Turning to possible heterogeneity in misreporting, Figure 3 shows the response biases characterizing the question on physical violence across respondent sub-groups.[18] We find that men with secondary education and those from high FGM prevalence ethnic groups significantly over-report physical violence by, respectively, 12.8 and 11.6 percentage points (see Table E1). It appears that men educated to secondary level are much less violent (17.4%) than men with lower levels of education (44.6%) as well as those from high FGM prevalence ethnic groups (22.6%) compared to those from low FGM-prevalence groups (37.7%), whereas these prevalence rates in DQ are similar regardless of the level of education, and higher among men from low FGM prevalence ethnic groups.

Figure 2: GBV practices: Prevalence rates and size of response biases



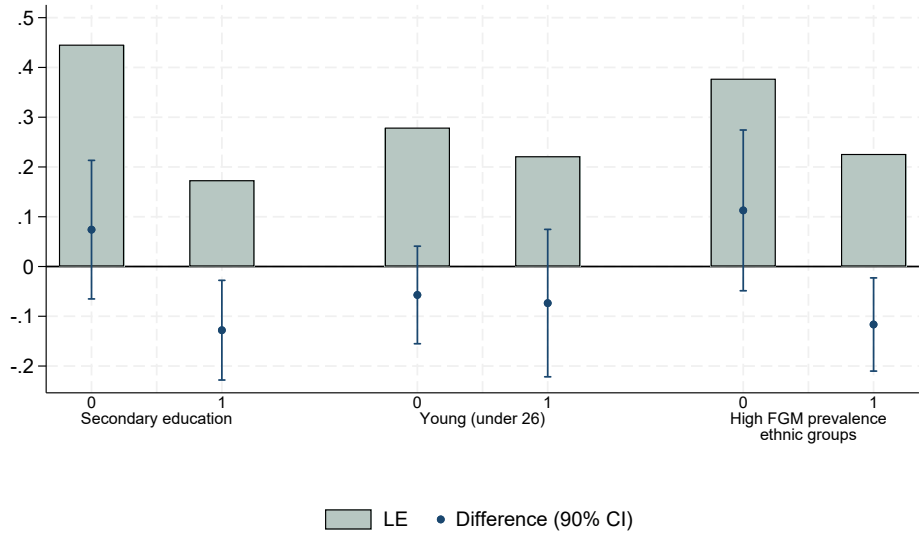
Finally, it is interesting to compare these results with those on IPV justification – keeping in mind that the phrasing of the two items is not fully symmetric. In fact, one relates to a man beating his wife in a specific situation; while the other relates, more generally, to a man

[16] Verbal violence against one's wife is excluded from our analysis given the list's likely violation of the no design effect assumption (see Section 1.3).

[17] Moreover, results reported in Appendix D, Table D1 show that our sample size, given the two survey techniques prevalence rates and baseline items' variance, is adequate for detecting biases for physical violence but not for rape.

[18] We do not investigate heterogeneity of response bias characterizing the question about rape, due to very low prevalence rates and potential design effects on rape by sub-groups (see footnote 11).

Figure 3: Perpetration of physical violence: Prevalence rates and size of response biases by sub-groups



hitting a woman, regardless of whether she is his wife and of the context. Still, the results show that violence is more justified (43.4%) than it is actually practiced (26.2%), suggesting that the justification of violence is not (only) an ex-post rationalization of a violent behavior. This is true in all subgroups, with men with a secondary education level being those who justify and practice it the least (respectively, 34.9% and 17.4%), and men without secondary education being those justifying it (60.8%) and practicing it the most (44.6%).

Several rationales, non-exclusive from each other, can be proposed in an effort to explain this discrepancy between support and perpetration. For instance, it could be that women, aware of men's opinions and afraid of being hit, behave in such a way so as to avoid being subjected to violence. Men's opinions would thus constrain women's behavior – eventually reducing the perpetration of violence. The LE-based prevalence rates that we obtain for perpetration of physical violence would then imply that men with secondary education interact with women who "behave" – hence they commit violence less often, while low-educated men interact with women who "misbehave" more often – hence they commit violence more frequently. While such an internalization of men's potential for violence by women (and especially their wives) is plausible, it would not explain why educated men significantly overreport their practice of violence. We would expect, by contrast, that a man in favor of IPV, but who does not in reality practice violence *because* women around him behave in accordance with his expectations, has no reason to declare that he actually does.

Another possible interpretation for the difference in prevalence between behaviors and opinions is linked to a change of social norms. There is qualitative evidence that changes in gendered social norms can be conducive to gaps between attitudes and behaviors, in a direction that is difficult to predict (Watson, 2014; Jewkes et al., 2015; Rost, 2021). Despite the absence of any legal sanctions against IPV today in Mali, civil society has conducted regular campaigns

against GBV in the past few years, also with the support of the United Nations Population Funds (UNFPA).[19] Our finding that low-educated men are more in favor of IPV than those with higher educational attainment, but also hide it more, could signal an ongoing evolution of norms. On the other hand, the overreporting of GBV practice by men with secondary education – who in fact practice GBV much less than their less-educated equivalents – could relate to persistent masculinity injunctions.

Other hypotheses could be drawn, including in connection with the slight difference in the phrasing of the two items, possibly relating them to social norms that do not fully overlap. Our empirical setting does not allow us to explore their plausibility. Still, we argue that the result suggesting that violence is more justified than practiced paves the way for further research dedicated to understanding how social norms dynamics can shape opinions and actual behaviors differently.

4 Robustness checks

4.1 Alternative specifications

We test whether our results are robust to not clustering the standard errors at the enumeration area level. We also run an alternative specification where we control only for enumerators' fixed effects, while keeping the standard errors clustered at the enumeration area level. In both cases our results are robust in terms of prevalence rates and response biases, as displayed in Table F1 in the Appendix.

4.2 Survey participation and refusal to answer

The team of enumerators visited 1,288 households, among which 86 refused to let the enumerator in, while 2 let him in but eventually did not consent to participate. Survey participation thus represents approximately 93% of the overall contacts made. Table 6 explores whether participation (defined as a dummy equal to 1 for the 1,200 interviewed households and to 0 for the 88 households who did not answer the survey) correlates significantly with the treatment status, defined before an enumerator knocks on a door. Reassuringly, participation is not significantly different between the treated and control groups.

Table 6: Survey participation across treated and control groups

	Treated (Std. Err.)	Control (Std. Err.)	Diff. (Std. Err.)
Survey participation	0.929 (0.010)	0.935 (0.010)	0.006 (0.014)
#	646	642	1,288

Besides survey participation, some respondents from the control group completed their questionnaire but refused to answer one or more sensitive direct question(s). Table 7 presents the

[19] For instance see https://mali.unfpa.org/sites/default/files/pub-pdf/newsletter_16_jours_contre_vff_2021.pdf (consulted 2024/01/17)

Table 7: Number and share of respondents in the control group who refused to answer each direct question

	Physical violence	Rape	Husband's control	Child marriage	FGM wife	FGM daughter	IPV justified
#	3	2	3	9	44	31	6
%	0.50	0.33	0.50	1.50	7.33	5.17	1.00

number of refusals to answer to each of the seven GBV-related questions we analyse. In four out of seven cases, only 2 or 3 individuals refused to answer. Respectively 6 and 9 individuals (i.e., 1% and 1.5% of respondents) refused to provide an answer to questions regarding the justification of IPV and child marriage. Questions about wife or daughter FGM report higher – but still low – rates of refusal to answer: respectively, 7.4% and 5.2%.

We further explore whether refusal to answer to questions on child marriage and FGM of wife and daughter is systematically correlated with individual characteristics.[20] For each of these three questions, we estimate a linear probability model of refusal to answer on an extensive set of individual variables and enumerator fixed effects.

Results are displayed in Table F2 in the Appendix. Individual characteristics together with enumerator fixed effects explain very little of the variance in the propensity to refuse to answer to these three questions, which should appease concerns about a possible non-randomness of refusal to answer. Still, older respondents and respondents with secondary education appear slightly more likely to refuse to answer the FGM-related questions, while respondents from high FGM prevalence groups are more reluctant to answer the question on child marriage. On the other hand, household heads are less likely to refuse to answer the two questions on FGM compared to their sons and, as far as daughter FGM is concerned, to other unrelated household members.

To further explore whether and how refusal to answer can affect our measure of response biases, we re-estimate the LE-based prevalence rate of each sensitive item (and the associated response bias) by dropping those respondents who refused to answer the corresponding direct sensitive question. The results of this test are displayed in column 2 of Table 8, while column 1 reproduces the baseline results. Reassuringly, the estimated bias is very stable in size and statistical significance – though slightly mitigated as far as question 8 is concerned.

The rest of Table 8 successively imputes various hypothetical DQ answers to those respondents who refused to answer to each sensitive question. In particular, we assume that, had they answered, they would all have answered "Yes" (column 3), all have answered "No" (column 5), or they would have answered "Yes" in a proportion in line with the results of our list experiment (column 4). While relying on arbitrary hypotheses, these three exercises aim to illustrate the extent to which the response bias we measure is related to refusal to answer standard direct questions. Together with the results displayed in Column 2, they suggest that in this study it is mostly untruthful answers (as opposed to refusal to answer) that explain misreporting in direct questions.

[20] Too few respondents refused to answer the other questions for us to run the same exercise on these questions.

Table 8: Estimated bias across various hypotheses on respondents who refused to answer a direct question

	Baseline	Refusals to answer in DQ...			
		...Missing in LE	...Yes in DQ	...LE-based prev. rate in DQ	...No in DQ
	(1)	(2)	(3)	(4)	(5)
Physical violence	-0.063	-0.064	-0.066	-0.063	-0.061
p-value	(0.212)	(0.206)	(0.190)	(0.214)	(0.224)
Rape	-0.007	-0.009	-0.010	-0.007	-0.007
p-value	(0.880)	(0.839)	(0.825)	(0.880)	(0.881)
Husband's control	-0.008	-0.011	-0.012	-0.008	-0.007
p-value	(0.852)	(0.802)	(0.792)	(0.852)	(0.880)
Child marriage	-0.069	-0.068	-0.078	-0.068	-0.063
p-value	(0.209)	(0.211)	(0.155)	(0.215)	(0.248)
FGM wife	-0.245	-0.261	-0.268	-0.227	-0.195
p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
FGM daughter	-0.207	-0.209	-0.220	-0.196	-0.168
p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
IPV justified	0.100	0.091	0.094	0.099	0.104
p-value	(0.048)	(0.072)	(0.065)	(0.050)	(0.040)

Notes: OLS estimations; LE equations include enumerator and age fixed effects, as well as dummies denoting secondary education level, belonging to an ethnic group characterized by a high prevalence of FGM, and the relationship with the household head.
Robust standard errors in parentheses (adjusted for clustering at enumeration area level); *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

5 Conclusion

According to the 2024 global Gender Inequality Index, Sub-Saharan Africa ranks among the lowest in the world (UNDP, 2024). Mali, in particular, ranks 172^{nd} out of 181 countries. Biased gender social norms are among the major obstacles in achieving gender equality and empowering all women and girls (SDG 5). Norms play an important role by influencing patterns of violence against women, as when individuals believe that violence is acceptable they might also directly enforce it. In this respect, men's opinions, who are most often the perpetrators of GBV, are key in the fight for gender equality, as their attitudes and behaviors may significantly shape the practice and acceptance of gender-based violence.

This paper studies men's opinions and behaviors related to GBV in Mali. We conduct a list experiment (LE) among a representative sample of 1,200 men living in Bamako to measure their practices and support of an extensive set of GBV dimensions (wife-beating, rape, child marriage, husband's control of wife's spending and FGM). By randomly assigning respondents to elicit answers to the sensitive questions using a DQ or LE technique, we estimate unbiased prevalence rates, i.e., free of response bias, and identify the extent to which views and behaviors are misreported with the direct question survey method.

Indirect questioning reveals that violent behaviors towards women as well as justification

of GBV are very widespread among men in Bamako. About one-quarter of men have already beaten a woman, nearly a third is in favor of husbands exercising financial control and female child marriage, while almost half justify IPV, think that a good wife is an excised woman and more than 53% would have their daughter circumcised if they had one. Moreover, the share of respondents who support IPV is larger than the share of those engaging in its practice: the list experiment shows that wife-beating is more justified (43.4%) than it is practiced (26.2%).

Interestingly, our results suggest that GBV practices related to women's marriage (child marriage and FGM) are still widely approved, but at varying degrees. The relatively low support for female child marriage we find suggests that a reform raising girl's legal age of marriage at 18 would probably encounter little opposition among the male population. In contrast, the importance of keep practicing female genital cutting seems to persist. These marriage-related practices are deeply connected to each other as delaying the age of marriage could increase girls' education, possibly substituting for the need of pursuing FGM with the aim of a positive marriage match, in line with García-Hombrados & Salgado (2023).

Our study also shows that men are not reluctant to openly state that they have already hit or raped a woman. Indeed, our results do not highlight declarative bias in responses to direct questions on the practice of GBV. However, we find substantial response biases concerning respondents' justification of intimate partner violence (IPV) and female genital mutilation (FGM), specifically when directly questioned men conceal justification of IPV but overstate their support for FGM. Response biases on support to FGM are uniformly distributed across the population, regardless of respondents' education level, age or ethnic group. We find, however, that men without secondary education significantly underreport their support for husband's control of wife's monetary resources and for IPV. Results thus call for caution with regards to the use of data collected using direct questions.

An important takeaway from our results is that support for discriminatory practices such as FGM is much lower than commonly thought, despite individuals preferring to conceal their true opinions. Conformity to the perceived norm may explain the persistence of this practice. Future research should further investigate the reasons for this considerable response bias, whether a "misperception" of social norms related to FGM exists, and assess whether providing information about the unbiased support for FGM would succeed in changing people's opinions and behaviors. Moreover, amid the slow decline in FGM rates across the whole of Africa (Kandala et al., 2018), the response bias we find might suggest that female genital cutting is a social norm that is likely evolving in Mali, possibly leaving room for policy reform.

References

28toomany (2018). *The law and FGM, an overview of 28 African countries.* Technical report, 28toomany and Thomson Reuters Foundation.

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

Alesina, A., Brioschi, B., & La Ferrara, E. (2021). Violence against women: A cross–cultural analysis for Africa. *Economica*, 88(349), 70 – 104.

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484), 1481–1495.

Asadullah, M. N., De Cao, E., Khatoon, F. Z., & Siddique, Z. (2021). Measuring gender attitudes using list experiments. *Journal of Population Economics*, 34(2), 367–400.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3), 491–507.

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

Blair, G. & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis*, 20(1), 47–77.

Bulte, E. & Lensink, R. (2019). Women's empowerment and domestic abuse: Experimental evidence from Vietnam. *European economic review*, 115, 172–191.

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. *American Economic Review*, 110(10), 2997–3029.

Camilotti, G. (2016). Interventions to stop female genital cutting and the evolution of the custom: Evidence on age at cutting in Senegal. *Journal of African Economies*, 25(1), 133–158.

Chuang, E., Dupas, P., Huillery, E., & Seban, J. (2021). Sex, lies, and measurement: Consistency tests for indirect response survey methods. *Journal of Development Economics*, 148, 102582.

Coffman, K. B., Coffman, L. C., & Ericson, K. M. M. (2017). The size of the lgbt population and the magnitude of antigay sentiment are substantially underestimated. *Management Science*, 63(10), 3168–3186.

Cullen, C. (2022). Method matters: The underreporting of intimate partner violence. *The World Bank Economic Review*, 37(1), 49–73.

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

Diabate, I. & Mesple-Somps, S. (2019). Female genital mutilation and migration in Mali: do return migrants transfer social norms? *Journal of Population Economics*, 32(4), 1125–1170.

Doyle, K., Levtov, R. G., Barker, G., Bastian, G. G., Bingenheimer, J. B., Kazimbaya, S., Nzabonimpa, A., Pulerwitz, J., Sayinzoga, F., Sharma, V., et al. (2018). Gender-transformative bandebereho couples' intervention to promote male engagement in reproductive and maternal health and violence prevention in rwanda: Findings from a randomized controlled trial. *PloS one*, 13(4), e0192756.

García-Hombrados, J. & Salgado, E. (2023). Legal bans, female genital cutting, and education: Evidence from Senegal. *The World Bank Economic Review*, 37(1), 74–92.

Gibson, M. A., Gurmu, E., Cobo, B., Rueda, M. M., & Scott, I. M. (2018). Indirect questioning method reveals hidden support for female genital cutting in South Central Ethiopia. *PloS one*, 13(5), e0193985.

Gibson, M. A., Gurmu, E., Cobo, B., Rueda, M. M., & Scott, I. M. (2022). Measuring hidden support for physical intimate partner violence: a list randomization experiment in South Central Ethiopia. *Journal of Interpersonal Violence*, 37(7-8), NP4238–NP4257.

Glynn, A. N. (2013). What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly*, 77(S1), 159–172.

Gulesci, S., Jindani, S., La Ferrara, E., Smerdon, D., Sulaiman, M., & Young, H. (2023). A stepping stone approach to norm transitions. (RWP23-013).

Gulesci, S., Leone, M., & Zafar, S. (2024). Domestic violence laws and social norms: Evidence from Pakistan.

Holbrook, A. L. & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly*, 74(1), 37–67.

Hossain, M., Zimmerman, C., Kiss, L., Abramsky, T., Kone, D., Bakayoko-Topolska, M., Annan, J., Lehmann, H., & Watts, C. (2014). Working with men to prevent intimate partner violence in a conflict-affected setting: a pilot cluster randomized controlled trial in rural côte d'ivoire. *BMC public health*, 14, 1–13.

Jamison, J. C., Karlan, D., & Raffler, P. (2013). *Mixed method evaluation of a passive mHealth sexual information texting service in Uganda*. Technical report, National Bureau of Economic Research.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual Reviews of Economics*, 7(1), 63–88.

Jewkes, R., Flood, M., & Lang, J. (2015). From work with men and boys to changes of social norms and reduction of inequities in gender relations: a conceptual shift in prevention of violence against women and girls. *The Lancet*, 385(9977), 1580–1589.

Joseph, G., Javaid, S. U., Andres, L. A., Chellaraj, G., Solotaroff, J., & Rajan, S. I. (2017). Underreporting of gender-based violence in Kerala, India: An application of the list randomization method. *World Bank Policy Research Working Paper*, (8044).

Jouvin, M. (2023). Addressing social desirability bias when measuring child labor use: An application to cocoa farms in Côte d'Ivoire. *The World Bank Economic Review*, (pp. lhad030).

Kandala, N.-B., Ezejimofor, M. C., Uthman, O. A., & Komba, P. (2018). Secular trends in the prevalence of female genital mutilation/cutting among girls: a systematic analysis. *BMJ global health*, 3(5), e000549.

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

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

Phillips, A. E., Gomez, G. B., Boily, M.-C., & Garnett, G. P. (2010). A systematic review and meta-analysis of quantitative interviewing tools to investigate self-reported hiv and sti associated behaviours in low-and middle-income countries. *International journal of epidemiology*, 39(6), 1541–1555.

Plummer, M. L., Wight, D., Ross, D. A., Balira, R., Anemona, A., Todd, J., Salamba, Z., Obasi, A. I., Grosskurth, H., Changalunga, J., et al. (2004). Asking semi-literate adolescents about sexual behaviour: the validity of assisted self-completion questionnaire (ASCQ) data in rural Tanzania. *Tropical Medicine & International Health*, 9(6), 737–754.

Poyker, M. (2023). Regime stability and the persistence of traditional practices. *Review of Economics and Statistics*, 105(5), 1175–1190.

Rosenfeld, B., Imai, K., & Shapiro, J. N. (2016). An empirical validation study of popular survey methodologies for sensitive questions. *American Journal of Political Science*, 60(3), 783–802.

Rost, L. A. (2021). 'By sharing work we are moving forward':change in social norms around men's participation in unpaid care work in Northern Uganda. *Oxford Development Studies*, 49(1).

Sanin, D. (2021). Do domestic violence laws protect women from domestic violence? evidence from Rwanda. *Georgetown University Initiative on Innovation, Development and Evaluation Working Paper*.

Tourangeau, R. & Yan, T. (2007). Sensitive questions in surveys. *Psychological bulletin*, 133(5), 859.

Traunmüller, R., Kijewski, S., & Freitag, M. (2019). The silent victims of sexual violence during war: Evidence from a list experiment in Sri Lanka. *Journal of Conflict Resolution*, 63(9), 2015–2042.

Treibich, C. & Lépine, A. (2019). Estimating misreporting in condom use and its determinants among sex workers: Evidence from the list randomisation method. *Health Economics*, 28(1), 144–160.

UNDP (2024). *Human Development Report 2023-24: Breaking the gridlock: Reimagining cooperation in a polarized world.* Technical report, United Nations Development Programme, New York.

UNSTAT (2015). *World's Women 2015: Trends and Statistics.* Technical report, United Nations.

UNSTAT (2020). *World's Women 2020: Trends and Statistics.* Technical report, United Nations.

Vaillant, J., Koussoubé, E., Roth, D., Pierotti, R., Hossain, M., & Falb, K. L. (2020). Engaging men to transform inequitable gender attitudes and prevent intimate partner violence: a cluster randomised controlled trial in North and South Kivu, Democratic Republic of Congo. *BMJ Global Health*, 5(5).

Watson, C. (2014). *Understanding changing social norms and practices around girls' education and marriage.* Technical report, Overseas Development Institute, London.

6 Appendix

A Lists of items

Table A1: Lists and questions about gender-based violence behaviors and opinions.

	Treatment group List (LE)	Control group Direct questions (DQ)
1.	*If married:* - I rarely get angry. - It is the man who should have the last word at home. *- Sometimes my words make my wife cry.* - Women should be able to access the same jobs as men.	*Do your words ever make your wife cry?*
2.	- My father always helped me when I needed it. *- I have already hit one or more women.* - Men should comfort babies as much as women do. - A woman should always go out accompanied.	*Have you ever hit one or more women?*
3.	- I know people who are in favor of abortion. - No one ever divorced in my family. - A man and a woman should decide together about their contraception. *- I have already forced a woman to have sex with me.*	*Have you ever forced a woman to have sex with you (although she didn't want to)?*
4.	- My neighbors have financial problems - It's up to the women to do the housework *- It's up to the man to decide how his wife's money should be spent* - I know people with HIV	*Is it up to the husband to decide how his wife's money should be spent?*
5.	- I trust people in my neighborhood - I prefer to eat apart from the women in my household *- It's better for a girl to get married before the age of 18* - If I were to get married today, I would choose a woman educated to the same level as me	*Is it better for a girl to get married before the age of 18?*
6.	- I recently attended a relative's wedding - We often argue in my family *- A good wife is an excised wife* - I trust women in politics	*Is a good wife an excised wife?*
7.	- Boys should participate in household chores *- If I were to have a daughter today, I would have her circumcised* - I know one or more unfaithful men -I know one or more people who had Covid	*If you had a daughter today, would you have her circumcised?*
8.	- My parents always got along - In the evening, women are not safe in the street - Corporal punishment should never be used in school *- A wife who does not respect her husband deserves to be hit*	*Does a wife who doesn't respect her husband deserve to be hit?*

B Pilot

We conducted a pilot survey to explore the response variation to possible baseline items and combine these items appropriately. The validity of LE rests on the assumptions of no design, ceiling and floor effects which greatly depend on the choice of baseline items (Blair & Imai, 2012; Glynn, 2013). Baseline items should neither be too common or too uncommon, nor too innocuous or too sensitive (Chuang et al., 2021). Furthermore, at least one baseline item should be negatively correlated to at least another baseline item and to the sensitive item.

Respondents to the pilot survey were randomly allocated to one type of questionnaire out of two. One included only direct questions (Pilot Questionnaire 1), while the other one included eight 4-items lists including the sensitive items (Pilot Questionnaire 2). In Questionnaire 1, all sensitive items along with a set of candidate baseline items were administrated under the form of a direct "Yes/No" question. Furthermore, answers provided to Questionnaire 2 allowed us to assess the acceptance and level of understanding of our lists.

The face-to-face pilot survey was conducted by GREAT (*Groupe de recherche en économie appliquée et théorique*) in Bamako in October 2022. Using the same sampling method as for the main survey, but in different enumeration areas, fifty adult men were interviewed. Among them, forty answered Questionnaire 1 and ten Questionnaire 2.

Table B1: Baseline items' descriptive statistics from the pilot survey

	Mean	sd.	Obs.
I1. Verbal violence			
Q1. Would you say that you rarely get angry?	0.69	0.468	39
Q2. Should men have the last word at home?	0.83	0.385	40
Q3. Should women be able to access the same jobs as men?	0.58	0.501	40
I2. Physical violence			
Q1. Did your father always help you when you needed it?	0.84	0.37	38
Q2. Should a woman always go out accompanied?	0.33	0.474	40
I3. Rape			
Q1. Do you know people in favor of abortion?	0.1	0.304	40
Q2. Should a man and a woman decide together about their contraception?	0.98	0.158	40
I4. Husband's control			
Q1. Do you know people with HIV?	0.18	0.385	40
I5. Child marriage			
Q1. Do you trust people in your neighborhood?	0.78	0.423	40
Q2. Do you prefer to eat apart from the women in your household?	0.53	0.506	40
Q3. If you were to get married today, would you choose a woman educated to the same level as you?	0.67	0.478	39
I6. FGM-wife			
Q1. Did you recently attend a relative's wedding?	0.8	0.405	40
Q2. Do you often argue in your household?	0.73	0.452	40
I7. FGM-daughter			
Q1. Should boys participate in household chores?	0.68	0.474	40
Q2. Do you know one or more unfaithful men?	0.51	0.506	39
Q3. Do you know one or more people who had Covid-19?	0.23	0.423	40
I8. IPV justified			
Q1. Did your parents always get along?	0.8	0.405	40
Q2. Are women unsafe on the street in the evening?	0.6	0.496	40
Q3. Do you think corporal punishment should never be used in school?	0.6	0.496	40

Table B2: Baseline items' correlations from the pilot survey

	I1		I2	I3	I5		I6	I7		I8	
	Q1	Q2	Q1	Q1	Q1	Q2	Q1	Q1	Q2	Q1	Q2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Q2	-0.0268		-0.228	0.103**	-0.0276		0.476***	-0.292**		-0.229*	
	(0.194)		(0.143)	(0.0498)	(0.135)		(0.161)	(0.143)		(0.114)	
Q3	0.00815	0.207			0	-0.0385		-0.297	0.344**	-0.0208	-0.146
	(0.154)	(0.128)			(0.147)	(0.174)		(0.188)	(0.170)	(0.131)	(0.158)

Notes: Robust standard errors in brackets; *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

C Exploring possible design effect

Table C1: Average number of items and share of respondents agreeing with j items, by list and treatment status.

	I1	I2	I3	I4	I5	I6	I7	I8
Zero items (T)	2.27	1.00	11.00	4.33	3.00	7.33	8.00	1.50
Zero items (C)	0.55	0.50	8.67	3.33	2.00	12.67	20.00	3.33
One item (T)	19.55	15.00	44.17	24.00	20.67	27.83	32.33	17.67
One item (C)	32.24	21.83	46.17	32.17	29.83	39.67	46.67	25.50
Two items (T)	35.98	41.00	34.33	44.50	40.67	37.50	40.50	36.67
Two items (C)	47.27	43.67	38.83	61.17	49.33	36.50	28.17	47.33
Three items (T)	30.31	32.83	10.00	25.67	27.17	22.50	16.17	33.83
Three items (C)	19.95	34.00	6.33	3.33	18.83	11.17	5.17	23.83
Four items (T)	11.90	10.17	0.50	1.50	8.50	4.83	3.00	10.33

Table C2: Test for design effect.

	Pr(R=r, S=0)	Pr(R=r, S=1)
I1. Verbal violence		
P-value	1.000	0.025
Bonferroni P-value	1.000	0.051
I2. Physical violence		
P-value	1.000	0.158
Bonferroni P-value	1.000	0.315
I3. Rape		
P-value	1.000	0.167
Bonferroni P-value	1.000	0.335
I4. Husband's control		
P-value	1.000	0.183
Bonferroni P-value	1.000	0.367
I5. Child marriage		
P-value	1.000	0.134
Bonferroni P-value	1.000	0.267
I6. FGM wife		
P-value	1.000	1.000
Bonferroni P-value	1.000	1.000
I7. FGM daughter		
P-value	1.000	1.000
Bonferroni P-value	1.000	1.000
I8. IPV justified		
P-value	1.000	1.000
Bonferroni P-value	1.000	1.000

D Bias-variance trade-off

The main advantage of using list experiments is to produce prevalence rates closer to the real prevalence of sensitive behaviors than those obtained with declarative data. However, while reducing response bias, LEs have a lower statistical precision. Direct questions may be biased, but they produce low-variance estimates. As Blair et al. (2020) illustrate, the gain in using list experiments compared to direct questions is significant when either the sample size or the bias is large enough (above 1,000 observations or above 5.5 percentage points). With a sample of 2,000 individuals the bias must be greater than four percentage points.

Following Blair et al. (2020) and Lépine et al. (2020), we estimate the minimum sample size for which our list experiments are likely to produce more valid results than the direct question. This corresponds to the sample size required to ensure that the list experiment method has a lower root mean square error than the direct question method given the observed bias (*Bias*) and the estimated variance of the list experiment in the control group (*Var(0)*). We then compare this minimum sample size to our survey sample size to assess whether our sample is large enough to prefer the list experiment to the direct question technique.

Results reported in Table D1 show that the survey sample size is large enough to obtain valid estimates with the list experiments for five items out of eight. The list experiment is preferred to the direct question technique in the presence of a response bias of more than 4.5 percentage points (which is the magnitude of the bias for item 1). This is consistent with the results from Blair et al. (2020) meta-analysis with a sample size similar to ours. In turn, our sample size is too small to detect the response bias associated with items 3 and 4, while it is almost able to detect the bias of item 1.

Table D1: Bias-variance trade-off and sample size

	List experiment			*Direct question*		**Bias**	**Min. sample size**	**Survey sample size**
	Var(0)	***b***	***s.e.***	***b***	***s.e.***			
I1	0.527	0.435	0.065	0.479	0.026	-0.045	1046	719
I2	0.567	0.251	0.046	0.325	0.019	-0.074	403	1200
I3	0.546	0.019	0.045	0.022	0.006	-0.003	211985	1200
I4	0.363	0.315	0.042	0.325	0.019	-0.010	15115	1200
I5	0.545	0.325	0.049	0.382	0.020	-0.058	644	1200
I6	0.726	0.433	0.053	0.680	0.020	-0.247	46	1200
I7	0.655	0.554	0.049	0.745	0.018	-0.192	71	1200
I8	0.621	0.420	0.048	0.333	0.019	0.087	328	1200

Notes: Var(0) corresponds to the variance of the list experiment for the control group. *b* and *s.e.* are the prevalence rates and associated standard errors of the list experiment and the sensitive direct questions. *Bias* is the size of the response bias affecting the sensitive direct questions. *Min. sample size* is the minimum sample size required for the mean-squared error of the list experiment to be lower than that of the direct question. *Survey sample size* is the sample size used in the list experiments for each item.

E GBV practices

Table E1: GBV practices: prevalence rates and response biases.

Panel A: average prevalence rates and response biases			
		I2 – Hit	**I3 – Rape**
	LE	0.262***	0.015
		(0.050)	(0.047)
	DQ	0.325***	0.022***
		(0.018)	(0.006)
	LE-DQ	-0.063	-0.007
	P-value	0.212	0.880
	N LE	1200	1200
	N	597	598

Panel B: group-specific prevalence rates and response biases			
		I2– Hit	
		0	**1**
Secondary education	LE	0.446***	0.174**
		(0.085)	(0.060)
	DQ	0.372***	0.302***
		(0.034)	(0.023)
	LE-DQ	0.074	-0.128
	P-value	0.382	0.061
Young	LE	0.279***	0.222*
		(0.055)	(0.085)
	DQ	0.336***	0.295***
		(0.023)	(0.035)
	LE-DQ	-0.057	-0.073
	P-value	0.337	0.414
High FGM prevalence ethnicity	LE	0.377***	0.226***
		(0.101)	(0.058)
	DQ	0.265***	0.343***
		(0.038)	(0.022)
	LE-DQ	0.113	-0.116
	P-value	0.251	0.041

Notes: Panel B: Column headed 0 reports LE-based and the DQ-based prevalence rates for the subgroup of the omitted category ($\hat{\gamma}_k$ and $\hat{g}_k$ respectively), while column headed 1 reports the same prevalence rates for the non-omitted subgroup ($(\hat{\gamma}_k + \hat{\delta}_k)$ and $(\hat{g}_k + \hat{\mu}_k)$, respectively). LE-DQ of Panel B Column 0 reports the difference between the LE-based and the DQ-based prevalence rate for the subgroup of the omitted category $(\hat{\gamma}_k - \hat{g}_k)$, while column 1 reports the same difference for the non-omitted subgroup $((\hat{\gamma}_k + \hat{\delta}_k) - (\hat{g}_k + \hat{\mu}_k))$

Secondary education refers to the achievement of a secondary education level; Under 25 years old equals one if the respondent is under 25 years old; High FGM prevalence ethnicity equals one if the respondent belongs to a high FGM prevalence ethnic group.

OLS estimations; LE equations include enumerator and age fixed effects, as well as dummies denoting secondary education level, belonging to an ethnic group characterized by a high prevalence of FGM, and the relationship with the household head.

Robust standard errors clustered at the EA level in parentheses.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

F Robustness tests

Table F1: Robustness checks

	I2	I3	I4	I5	I6	I7	I8
Panel A. No cluster							
DQ	0.325	0.022	0.325	0.382	0.680	0.745	0.333
LE	0.260	0.016	0.318	0.313	0.437	0.541	0.434
Diff.	-0.065	-0.006	-0.007	-0.069	-0.243	-0.205	0.101
p-value	(0.195)	(0.898)	(0.873)	(0.172)	(0.000)	(0.000)	(0.052)
Panel B. No controls, cluster EA							
DQ	0.325	0.022	0.325	0.382	0.680	0.745	0.333
LE	0.251	0.019	0.315	0.325	0.433	0.554	0.420
Diff.	-0.074	-0.003	-0.010	-0.058	-0.247	-0.192	0.087
p-value	(0.142)	(0.944)	(0.835)	(0.296)	(0.000)	(0.000)	(0.080)
Panel C. Baseline specification							
DQ	0.325	0.022	0.325	0.382	0.680	0.745	0.333
LE	0.260	0.016	0.318	0.313	0.437	0.541	0.434
Diff.	-0.065	-0.006	-0.007	-0.069	-0.243	-0.205	0.101
p-value	(0.197)	(0.901)	(0.868)	(0.210)	(0.000)	(0.000)	(0.047)

Notes: OLS estimations.
Panel A: LE equations include enumerator and age fixed effects, as well as dummies denoting secondary education level, belonging to an ethnic group characterized by a high prevalence of FGM, and the relationship with the household head. Robust standard errors in parentheses.
Panel B: LE equations include enumerator fixed effects. Robust standard errors in parentheses (adjusted for clustering at enumeration area level).
Panel C: LE equations include enumerator and age fixed effects, as well as dummies denoting secondary education level, belonging to an ethnic group characterized by a high prevalence of FGM, and the relationship with the household head. Robust standard errors in parentheses (adjusted for clustering at enumeration area level).
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

Table F2: Determinants of refusal to answer in DQ

	All (control) respondents			Married (control) respondents		
	I5	I6	I7	I5	I6	I7
Age	-0.000	0.002*	0.001*	-0.001	0.002	0.001**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Secondary education achieved	0.014	0.034*	-0.003	0.004	0.050**	-0.002
	(0.01)	(0.02)	(0.02)	(0.01)	(0.03)	(0.02)
Ethnic group pro FGM	0.023***	-0.031	-0.017	0.021*	-0.032	-0.031
	(0.01)	(0.03)	(0.02)	(0.01)	(0.04)	(0.04)
HH head's father	-0.020	0.039	-0.002	-0.024	0.076	-0.027
	(0.02)	(0.07)	(0.03)	(0.02)	(0.15)	(0.03)
HH head's son	-0.014	0.073**	0.068*	-0.004	0.090	0.150**
	(0.02)	(0.04)	(0.04)	(0.03)	(0.06)	(0.06)
HH head's relative	0.009	0.027	0.033	-0.005	0.046	0.043
	(0.02)	(0.03)	(0.03)	(0.01)	(0.04)	(0.03)
HH member unrelated	-0.015	0.050	0.135	-0.036	0.267	0.651**
	(0.02)	(0.07)	(0.10)	(0.03)	(0.26)	(0.26)
At least one adult woman in HH	-0.027	0.007	0.001			
	(0.04)	(0.04)	(0.04)			
Married	0.002	0.027	0.025			
	(0.02)	(0.03)	(0.03)			
Spouse lives in HH				-0.020	0.007	0.029
				(0.03)	(0.07)	(0.05)
Constant	0.057	-0.048	-0.044	0.096	0.000	-0.044
	(0.04)	(0.06)	(0.05)	(0.06)	(0.09)	(0.07)
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	600	600	600	366	366	366
R^2	0.038	0.072	0.069	0.046	0.086	0.152
Adjusted R^2	0.005	0.040	0.037	-0.007	0.036	0.105

OLS estimations. Robust standard errors clustered at the enumeration section level in parentheses.
* p<0.10, ** p<0.05, *** p<0.01