Journal of Development Economics 176 (2025) 103479

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

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

DIAL-LEDa, Université Paris-Dauphine, Université PSL, CNRS, IRD, 75016 PARIS, France

ARTICLE INFO

Dataset link: https://doi.org/10.23708/A6GRZK

Keywords:
Gender-based violence
List experiment
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 of eight GBV-related opinions and behaviors among men in Bamako, the capital city, by administering a set of list experiments that we compare to a set of direct questions to estimate response biases. We find large support for GBV: nearly one respondent in two supports female genital mutilation or intimate partner violence. Besides, one in four has already physically hit an adult woman. Our results also show that several questions suffer from significant response biases when asked directly. Support for female genital mutilation is overreported, with actual approval being lower than openly stated. Conversely, justification of intimate partner violence is underreported, likely due to social pressure against it. While response bias varies little with respondent characteristics, prevalence rates are systematically lower among men with a secondary level of education. Our results are in line with response bias being shaped by the legal framework addressing GBV as well as prevailing social norms, highlighting the need for caution when using data collected through direct questioning.

1. 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 appear to be decreasing in many of the 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 support for GBV because of shame or fear of lack of confidentiality (Tourangeau and Yan, 2007). Therefore, standard survey data may not be accurate and may generate biased prevalence rates.[1]

Recent works highlight the role of male social norms for improving attitudes and behaviors towards women (Jewkes et al., 2015; Bursztyn et al., 2020; Vaillant et al., 2020), pointing towards the need to assess men's attitudes towards violence against women in the fight against GBV. Accounting for the men's perspective is also intended at facilitating the design, implementation and acceptance of anti-GBV policies and at encouraging the targeting of men and boys to prevent GBV.

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 child marriage among a representative sample of 1,200 adult males living in Bamako, the capital city of Mali. To limit response bias, we rely on a set of list experiments (LEs), a survey technique commonly used in social sciences to estimate arguably unbiased prevalence rates.[2] LE reduces respondents' reluctance

* Corresponding author.
E-mail addresses: olivia.bertelli@dauphine.psl.eu (O. Bertelli), thomas.calvo@ird.fr (T. Calvo), emmanuelle.lavallee@dauphine.psl.eu (E. Lavallée), marion.mercier@cnrs.fr (M. Mercier), sandrine.mesple-somps@ird.fr (S. Mesplé-Somps).

[1] For instance, research shows that in Ethiopia, where FGM has become illegal, the indirectly elicited support among both men and women for FGM is in fact higher than the one measured with standard direct question surveys (De Cao and Lutz, 2018; Gibson et al., 2018).

[2] For a recent review on the use of LEs in social sciences see Chuang et al. (2021).

https://doi.org/10.1016/j.jdeveco.2025.103479
Received 17 September 2024; Received in revised form 18 February 2025; Accepted 19 February 2025
Available online 17 April 2025

to reveal the truth by asking questions indirectly rather than directly, masking respondents' answers both to interviewers and researchers. We administer both LEs and Direct Questions (DQs) to provide evidence of response bias, that is the extent to which opinions and behaviors are concealed in DQs.

To our knowledge, this is the first analysis providing prevalence rates arguably net of response bias 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. In Mali, the legal age of marriage is 16 for women (vs. 18 for men) and no law against domestic violence or FGM exists, as attempts to pass such laws have systematically failed amid the opposition of Muslim national movements.

Our main results show that male 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 cut wife, 54% are in favor of having the genitalia of an hypothetical daughter cut, 43% justify IPV if a woman disrespects her husband. One third supports female child marriage and perceives as normal for husbands to control the spending of their wives' money. Men with secondary education are the ones supporting GBV the least, whereas other characteristics — such as age or ethnic group — do not seem to matter.

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 (−10.2 percentage points) and massively overreport their support for wife and daughter genital mutilation (respectively, +24.9 and +21.6 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. The only notable difference pertains to men with low levels of formal education who underreport support for IPV. The fact that support for FGM is about 30% lower in the LE than openly stated is even more striking given that FGM prevalence rate in Mali is among the highest in the world. It may signal room for policy reforms.

A second set of results concerns the perpetration of violence, for which we find similar prevalence rates with the two survey techniques. About 26% of respondents declare having ever physically hit an adult woman, while 1% report having ever raped a woman. While the lack of response bias is somehow surprising, it would be in line with a perceived general social acceptance of GBV, whereby perpetrators are not ashamed to tell the truth. Comparing these results with those pertaining to the justification of GBV reveals that wife-beating is, on average, more justified (44%) than practiced (26%) by men, suggesting that justification of violence is not an ex-post rationalization of violent behaviors.

Our work 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 and Krosnick, 2010; Rosenfeld et al., 2016), loan use (Karlan and 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 and Lépine, 2019; Lépine et al., 2020) and child labor (Jouvin, 2023).[3] More specifically, our work contributes to the set of studies aiming to identify unbiased GBV prevalence rates. A number of articles explore reporting biases about IPV — in particular, Joseph et al. (2017) in Kerala in India, Bulte and Lensink (2019) in Vietnam, Traunmüller et al. (2019) about sexual violence during the Sri Lankan civil war, Lépine et al. (2020) in rural Burkina Faso, Agüero and Frisancho (2022) in Lima's poor districts in Peru, Cullen (2022) in the Kebby state of Nigeria, and Gibson et al. (2022) in a South Central Ethiopian community. Support for child marriage is also found to be affected by misreporting in Bangladesh (Asadullah et al., 2021), as is the support for FGM in two different regions of Ethiopia (Gibson et al., 2018; De Cao and Lutz, 2018).

The novelty of our work lies in studying potential perpetrators rather than victims. Very few studies have so far included male respondents. We add to this scarce body of work by exploring men's attitudes and behaviors in a West African capital city, a context of stark gender inequalities. To the best of our knowledge, the first attempt to measure support for FGM by comparing answers to a standard direct question with a list experiment is the work by Gibson et al. (2018), in the Arsi Oromo region of Ethiopia, where both women and men are surveyed. The authors find that support for cutting daughters and daughters-in-law is generally underreported. Male respondents are also surveyed by Gibson et al. (2022) who investigate justification for domestic violence in the same Ethiopian region and find that men significantly underreport it by 17 percentage points. Finally, Cullen (2022) reports in the supplementary appendix the results of a list experiment ran in Rwanda surveying couples about domestic violence. Men, in particular, are asked about whether in the past twelve months they tried to limit their wife's contacts with her family and whether they have tried to hurt her or someone close to her. She finds evidence in line with a significant underreporting only for the first item by 31.8 percentage points.

This article also contributes to the debate about the heterogeneity in response bias as measured with list experiments. A number of studies highlight heterogeneity in misreporting of GBV-related items across education levels, though in opposite directions.[4] Joseph et al. (2017) also observe that the youngest and oldest respondents underreport domestic violence more than those of the middle-age cohort. We do not find systematic heterogeneity across respondent characteristics, with the exception of support for IPV that is more underreported by men with lower levels of formal education.

We also add to the recent debate about the capacity of LEs to measure 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 LEs' higher variance. Blair et al. (2020) highlight that bias in sensitive items in political science is often small — around 5 percentage points — limiting the contribution of LEs. We take part in this debate by showing that, when it comes to male support for GBV in Mali, 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, our study adds to the 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, leading to an overestimation of the true intervention's impact on men's self-reported attitudes. More broadly, our results call for caution in using declarative data about GBV-related attitudes.

The rest of the paper is structured as follows. Section 2 provides some insights on the Malian context, Section 3 presents the sample, the

[3] See Blair et al. (2020) for a recent review of LEs in political science.

[4] In particular, women's education is positively related to the underreporting of IPV in Kerala (Joseph et al., 2017), in Nigeria (Cullen, 2022) and among female micro-credit clients in Lima (Agüero and Frisancho, 2022), and support for FGM is mostly underreported by educated respondents in Ethiopia (Gibson et al., 2018). On the contrary, support for IPV is mostly underreported by uneducated adults in Ethiopia according to Gibson et al. (2022), and De Cao and Lutz (2018) observe that uneducated women are particularly prone to underreporting their support for FGM in the Afar region in Ethiopia.

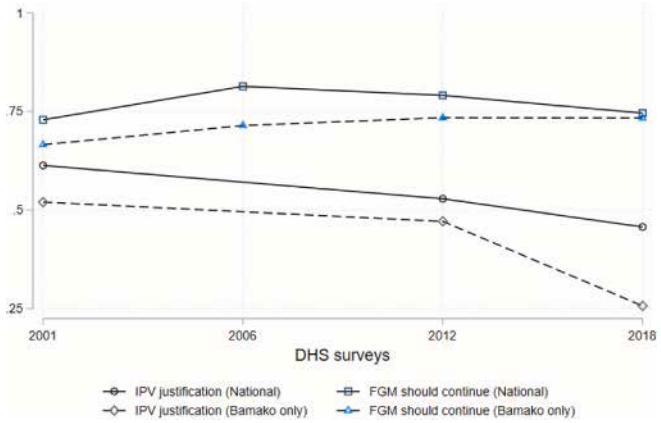

Fig. 1. Share of male respondents supporting IPV or FGM in Mali by DHS survey year.

randomization design, the lists of items under scrutiny and provides evidence supporting the validity of the experimental design. In Section 4 we outline the empirical specifications for data analyses. Section 5 reports the results. We first present the results on GBV justification and then examine whether misreporting GBV justification is homogeneous in our sample. Second, we turn to the results on GBV perpetration. Section 6 draws the main conclusions.

2. The Malian context

This study is carried out in Bamako, the capital city of Mali, one of the lowest ranked countries in the world in terms of the Human Development Index (188th out of 193 countries) and the Gender Inequality Index (172th out of 181 countries) (UNDP, 2024). According to the latest available Demographic and Health Survey (DHS, 2018), GBV is widespread in Mali. Almost one Malian woman out of two has been victim of GBV in her lifetime, 92% underwent FGM, and 60% were married during childhood (i.e., before the age of 18) — against 2% of Malian men. Discriminatory gender norms are also widely accepted as 80% of women justify wife-beating — the second highest rate in the world after Afghanistan (UNSTAT, 2020) — and 76% of women declare that FGM should continue.[5]

Nevertheless, the civil society — particularly Malian women's associations — and international organizations, as well as the Malian government, have taken action to fight GBV (Alcalde et al., 2015). In 1985, Mali ratified the Convention on the Elimination of All Forms of Discrimination against Women; and the Ministry for the Promotion of Women, Children and the Family was set up in 1997. The first national program to fight the practice of FGM was adopted in 2002; and for many years Mali has actively taken part in the International Day for the Elimination of Violence against Women, through national information and awareness-raising campaigns.

Yet, various initiatives undertaken by the government to legislate on intra-family relations have been subject to tensions with the country's Muslim authorities (Schulz, 2012). The reform of the family code adopted in 2009 by the parliament provoked strong opposition from the High Islamic Council of Mali (HCIM).[6] The resulting protests and demonstrations were successful enough to have the law changed. In December 2011, an amended version of the code incorporating the notion of a wife's duty of obedience to her husband, the legalization of religious marriage, the non-recognition of children born out of wedlock and the reduction of the female minimum age of marriage to 16 years old was adopted and promulgated, thus satisfying most of the HCIM demands (Koné and Calvès, 2021; Whitehouse, 2022).

Besides intra-family relations, FGM is also a major political issue in Mali, seized by Malian Muslim movements. Various draft laws to ban FGM have failed to gain approval and pass through parliament. Most recently, on the International Day of Zero Tolerance for FGM in February 2017, the Minister of Justice stated his intention to legally ban the practice of FGM by the end of the year. The leader of the HCIM immediately denounced this issue as being introduced by the West to foment disorder and ensure control over Mali, and threatened that the mobilization against such a law would spread throughout the country (28toomany, 2018; Amselle, 2018). The draft law was never submitted to the parliament, and no other law against any form of GBV has been on the political agenda since then.

Thus, while being characterized by high levels of GBV, Mali experienced both attempts to fight violent practices towards women and conservative resistance. In this context, looking at all available DHS surveying men provides a snapshot of the trend in Malian men's declared opinions about IPV and FGM. This exercise suggests the existence of contrasted patterns (Fig. 1). The share of men justifying IPV considerably decreased over time, from 61% in 2001 to 46% in 2018 (and from 52% to 26% in Bamako). This echoes the fact that, according to the 2022 Afrobarometer data, 74% of men consider that the government should do more to promote women's rights (Coulibaly et al., 2022). This decreasing trend is also consistent with the multiple information campaigns against GBV implemented in the past few years, and with the growing militancy in the civil society repeatedly calling for laws to punish the perpetrators of GBV.

These figures, however, contrast with those revealing a persistent support for FGM that was as high as 75% in 2018 (73% in Bamako). Available DHS data show that support for FGM, in fact, slightly increased since 2001 (73% in Mali and 67% in Bamako). While male justification of domestic violence seems to decline over time, possibly due to social pressure, the practice of FGM appears to be constantly supported across the country, possibly in line with the influence of conservative Islamist movements and/or deep-rooted traditions. Notwithstanding, these trends could signal either sincere attitudes and/or the internalization of perceived prevailing social norms — be they against IPV or in favor of FGM.

3. Survey experiment

3.1. 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 in 2022 by the National Statistics Office (INSTAT) during the mapping phase of the last General Population Census. Within each EA, we implemented a random walk protocol to select the sampling households. In each household, enumerators randomly selected one permanent adult male member speaking Bambara or French to be surveyed.[7]

Respondents were randomly assigned to the treatment or control group in the following way. Each enumerator was assigned to interview

[5] Note that DHS surveys do not provide information on opinions about child marriage.

[6] The HCIM was created in 2002 by the Malian government, with the aim of providing a single official interlocutor to the political authorities for all questions related to the practice of Islam.

[7] 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.

Table 1
Randomization balance checks.

	Control			Treatment			
	Obs. (1)	Mean (2)	S.D. (3)	Obs. (4)	Mean (5)	S.D. (6)	Diff. (7)
Age	600	37.76	15.99	600	36.76	15.41	−0.994
Under 26 years old	600	0.28	0.45	600	0.30	0.46	0.027
Primary education achieved	600	0.73	0.44	600	0.75	0.43	0.017
Secondary education achieved	600	0.42	0.49	600	0.45	0.50	0.036
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	0.95	0.21	600	0.95	0.21	−0.005
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							0.59
p-value							0.836

Notes: column (7) reports the estimated coefficient of the treatment dummy variable regressed against each characteristic reported in the first column, controlling for enumerators' fixed-effects. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

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 and a treatment one if the enumerator ID number was odd.[8] The difference between the two groups relied only on the types of lists administrated. The treatment group answered eight lists each one with a sensitive item. The control group was presented the same eight lists but without the sensitive items, that were instead administrated with a standard direct-question technique.

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 three quarters of respondents completed at least primary education while about 43% completed secondary education,[9] 60% are married and the vast majority live with an adult woman in the household. High FGM prevalence ethnic groups correspond to groups (defined based on respondent mother tongue) with rates of genital cutting close to 90% as documented in Diabate and Mesple-Somps (2019). They include (among others) the Bambara, Malinke, Fulani and Senufo.

Survey participation was high, above 93%. In Online Appendix F.2, we show that, reassuringly, participation does not significantly correlate with the treatment status. Moreover, few respondents refused to answer certain questions. In the same Appendix, we show that individual characteristics as well as enumerator fixed effects do not explain refusals to answer, which is reassuring regarding their potential non-randomness. Still, we explore the implications of refusals to answer and show that they barely affect our measures of response biases.

3.2. 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 how many of these statements they experienced or agree with, without specifying which ones. In the control group, lists are made of J so-called "baseline" items, while in the treatment group they are made 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 not differ across groups. The difference between the two groups average responses provides an estimate of the sensitive item prevalence 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, where 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.[10] Three items concern the perpetration of violence (verbal, physical, and sexual violence), while the other five are related to opinions about GBV (economic control, female child marriage, female genital cutting, and physical intimate partner violence). Their ranks in the lists were randomly determined.

It is important to mention that we did not randomize the order of lists across respondents, implying a possible risk of fatigue for the final lists.[11] To mitigate this risk to the fullest extent, we made the survey as short as possible — interviews were, in most cases, between 15 to 20 minutes long. Moreover, to encourage respondents to complete the questionnaire, we provided them with FCFA 1,000 of air time at the end of the interview. This was announced just after their consent to answer our survey was obtained.

This survey design 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 question administrated to the control group.

3.3. 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, and the hypothesis of no design effect. We follow Blair and 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 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.

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 (resulting in a large share of "No" answers) or too common (resulting 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 (see Online Appendix

[8] 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.

[9] Data from the representative Permanent and Modular household survey (EMOP) collected contemporaneously with ours by INSTAT indicate that 72.8% of adult men living in Bamako have primary education or more and 35.3% have secondary education or more.

[10] Table A.1 in the Appendix A also presents the non-sensitive items.

[11] Somewhat reassuringly, all respondents fully completed the questionnaire after giving their consent.

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 a cut wife	Is a good wife a cut wife?
I7. FGM daughter	If I were to have a daughter today, I would have her cut	If you had a daughter today, would you have her cut?
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?

B). A number of possible baseline items were administered under the form of direct "Yes/No" questions to identify high-prevalence and low-prevalence items, as well as possible sensitive items, more frequently refused to be answered by respondents when asked directly. We excluded items with more than 5% of refusals to answer or that were not generating enough variance, and designed the lists accounting for correlations between baseline items following the guidelines of Chuang et al. (2021). The pairwise correlations between items are not too large and, for most of the lists, are both positive and negative, limiting the risks of floor and ceiling effects (see Tables B1 and B2 in Online Appendix B).[12]

Following Glynn (2013), 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). Fig. 2 is a graphic representation of these shares by treatment status (see Table C1 in Online Appendix C for the precise proportions). Attention should be paid to the tails of the distributions in the control group, as a high share of zeros could signal floor effects while a high share of threes could signal ceiling effects.

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 whose true answer is 1 may declare 0, whereas those who would answer 0 may choose 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 with the sensitive item, thus yielding an upward bias in the acceptance of an hypothetical daughter genital mutilation 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 Online Appendix B).

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 our LE to underestimate the prevalence rate of physical violence.

The validity of a list experiment also rests on the assumption of no design effect. This arises when the inclusion of a sensitive item affects the answers to the baseline items. To explore the risk of design effect, we follow Glynn (2013) and 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 group shares would signal the presence of a design effect. 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 Table C2 (Online Appendix C). 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.[13]

4. Empirical strategy

The first empirical step of our analysis consists in measuring the prevalence rates with the two survey techniques for each sensitive item and 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:

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

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

The prevalence rate of sensitive item k using the LE technique, $\hat{\gamma}_k$, is measured by comparing the control and treatment groups in the following regression:

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

[12] 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 tested in 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.

[13] 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 26 years old in item 3. Results are available upon request.

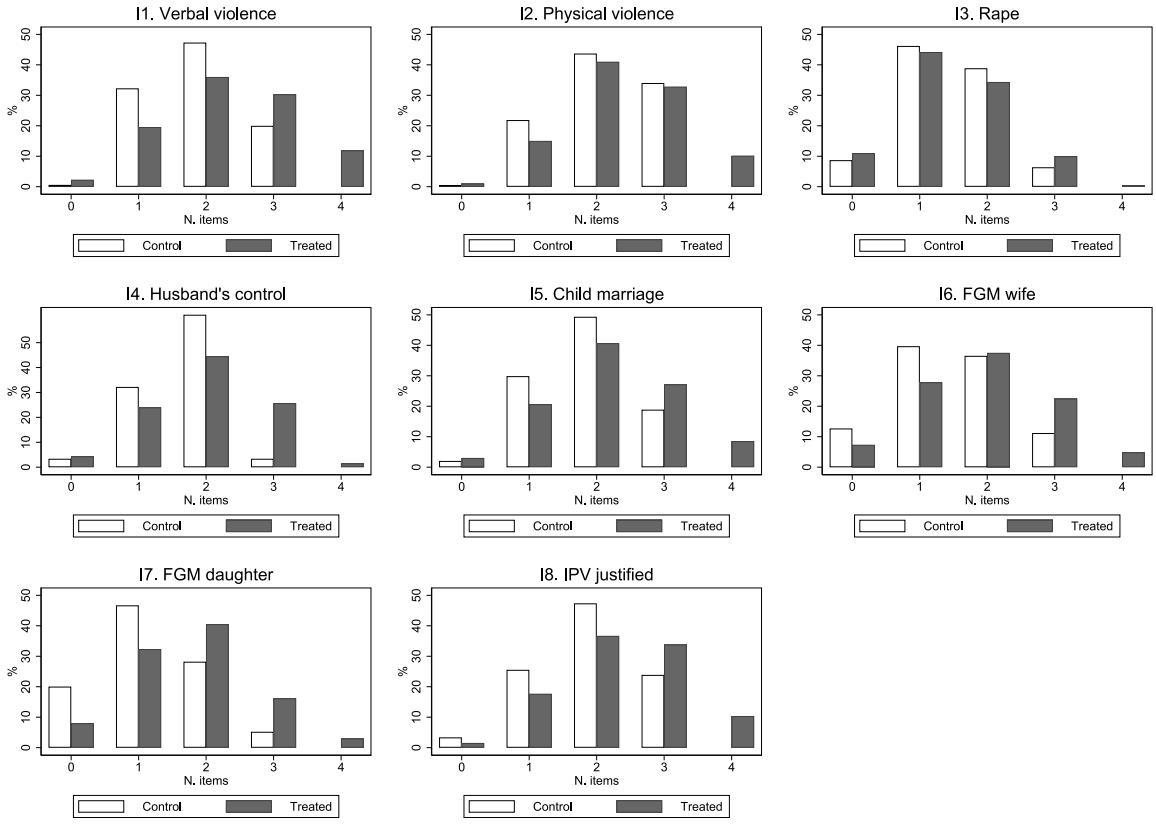

Fig. 2. Share of respondents agreeing with j items, by list and treatment status.

Table 3
Share of respondents answering at least j items by list and treatment status.

	I1	I2	I3	I4	I5	I6	I7	I8
At least one item (T)	97.73	99.00	89.00	95.67	97.00	92.67	92.00	98.50
At least one item (C)	99.45	99.50	91.33	96.67	98.00	87.33	80.00	96.67
Diff. T–C	−1.72*	−0.50	−2.33	−1.00	−1.00	5.33	12.00	1.83
At least two items (T)	78.19	84.00	44.83	71.67	76.33	64.83	59.67	80.83
At least two items (C)	67.21	77.67	45.17	64.50	68.17	47.67	33.33	71.17
Diff. T–C	10.97	6.33	−0.33	7.17	8.17	17.17	26.33	9.67
At least three items (T)	42.21	43.00	10.50	27.17	35.67	27.33	19.17	44.17
At least three items (C)	19.95	34.00	6.33	3.33	18.83	11.17	5.17	23.83
Diff. T–C	22.26	9.00	4.17	23.83	16.83	16.17	14.00	20.33

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

where $Y_{i,k}$ denotes the response given by individual i to the list of items containing k ($Y_{i,k}$ 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. Eq. (2) yields an accurate estimate of the prevalence of item k provided that the assignment to treatment is random. To increase the estimates' precision, 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.[14]

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, showing the value-added of using the LE-technique rather than the DQ one.

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 and Imai (2012), we compute prevalence rates by education level, age category (less than 26 years old or 26 and older), and ethnicity. We expect these characteristics to matter for prevalence rates and response biases. In fact, respondent education and age are heterogeneity dimensions highlighted by previous works (Joseph

[14] Table F1 in Online Appendix F shows that our results are robust to controlling only for enumerator fixed effects without any other control variable.

et al., 2017; Gibson et al., 2018; De Cao and Lutz, 2018; Agüero and Frisancho, 2022; Cullen, 2022), while ethnic group practice of FGM is expected to be relevant in the Malian context.[15] We differentiate between groups with high and low FGM prevalence rates based on Diabate and Mesple-Somps (2019). Based on data from 2009, they observe a rate of FGM among girls below the age of 14 close to or even larger than 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 26 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 Eqs. (3) and (4).

In all our estimations, we cluster standard errors at the enumeration area-level.[16]

5. Results

5.1. 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 Eq. (1) and $\hat{\gamma}_k$ estimated from Eq. (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.

The LE prevalence rates reveal that a large share of respondents justify GBV practices. About 53% of men are in favor of FGM for an hypothetical daughter, about 43% consider that a good wife is a cut wife, and 43.6% justify IPV if a wife disrespects her husband. The dimensions that encounter the least support are female child marriage (31.6%) and husband's control of his wife's monetary resources (32%). These results show a relatively good acceptance of women's economic independence and a relatively weak support for female child marriage, whereas support for FGM and IPV appears to be widespread.

Some of these prevalence rates are subject to a declarative bias when measured with direct questions, in particular, those for the two items concerning the support for FGM and the item concerning IPV justification. When asked directly, more than two-thirds of respondents approve FGM: 68% declare that a good wife is a cut wife and 74.5% report that, if they had a girl, they would have her cut. These prevalence rates are consistent with those found in the latest available DHS data for Mali (2018) where 73% 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 24.9 and 21.6 percentage points, respectively, corresponding to 37% and 29% of the respective DQ-prevalence rates.[17]

This considerable misreport provides a sense of the sensitivity of FGM among the male population of Bamako, and suggests that openly expressing reservations about FGM practices is problematic. The Malian religious and institutional context likely explains the opposite results we find compared to De Cao and Lutz (2018) and Gibson et al. (2018), who show 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, widely practiced, and supported by Muslim organizations, support for FGM is overreported in DQ. This misreporting might, at least partly be explained by 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 IPV justification, our results also emphasize a significant response bias. When asked directly, 33% of men justify IPV when a woman disrespects her husband. This figure is close to those provided by the DHS data (2018) where 26% of men in Bamako (31% of men in urban areas) justify wife beating. However, a significantly larger share of men appears to justify IPV according to the LE. This proportion reaches nearly 44% in the LE, revealing a statistically significant 10.2 percentage point bias that corresponds to one third 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 highlight an underreporting of IPV justification among men in Ethiopia — though statistically significant only for men without secondary education. In Bamako, such underreporting can be interpreted in light of the decreasing trend in IPV justification by men observed in the DHS data, potentially reflecting greater social pressure against domestic violence.[18]

Overall, opinions about FGM and IPV appear to be sensitive subjects, leading to biased responses when directly asked. This calls for exercising caution with regard to using data collected with standard direct technique.

Finally, neither the support for husband's economic control nor for early marriage appear to be affected by response biases. This suggests that revealing information about these personal opinions is not perceived as sensitive.[19] As far as support for child marriage is concerned, the absence of response bias differs from the underreporting found among women in Bangladesh (Asadullah et al., 2021), a country where child marriage is banned, contrary to Mali where girls are legally authorized to get married starting from the age of 16.

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, LEs also increase the variance of estimates.

[15] Following the pre-analysis plan, we also explore possible heterogeneity depending on whether the respondent is the household head, completed primary school or not, is married or not and, as far as married respondents are concerned, whether he is in a polygamous or monogamous union. These dimensions do not appear to matter for heterogeneity of neither the prevalence rates nor the response biases. We do not have the information on whether respondents have at least one daughter.

[16] All our results are robust to not clustering the standard errors (results available upon request).

[17] 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 by our protocol.

[18] Men's IPV justification drops from 52% in 2001 to 26% in 2018 in Bamako (Fig. 1.)

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

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	0.320***	0.316***	0.431***	0.530***	0.436***
	(0.042)	(0.054)	(0.054)	(0.050)	(0.051)
DQ	0.325***	0.382***	0.680***	0.745***	0.333***
	(0.020)	(0.020)	(0.018)	(0.019)	(0.019)
LE-DQ	−0.005	−0.067	−0.249	−0.216	0.102
p-value	0.912	0.220	0.000	0.000	0.043
Sharpened *q*-value	0.544	0.282	0.001	0.001	0.094
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); *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

Following Blair et al. (2020) and Lépine et al. (2020), we estimate the minimum sample size for which our LEs are likely to produce more valid results than DQs. Results are reported in Online Appendix D, Table D1. They 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.

5.2. 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, that we measure as dummy variables. Specifically, we distinguish men with at least secondary level of education from those with a lower level of education; men aged under 26 years old from those older; and men belonging to an ethnic group with high vs. low FGM prevalence. 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 4, the analysis relies on the comparison between the results of the estimation of Eqs. (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 LEs reveals a number of interesting patterns (columns 1–3 of Table 5). As highlighted above, FGM is the form of GBV that is most widely approved, and this is generally true across respondent characteristics.

The widest prevalence gap is observed between men with and without secondary education. In turn, young and older men, or those from high- and low-FGM prevalence ethnic groups, exhibit similar levels. In particular, men with secondary education are much less likely to be in favor of husbands controlling their wife's money (16.1%, against 44.3% of men without secondary education), to prefer a genitally mutilated wife (24.0% against 57.8%), to support FGM for an hypothetical daughter (39.8% against 63.2%) or to justify IPV (32.6% against 52.1%). Support for child marriage is also lower among men educated to secondary level, though the difference is not statistically significant. Men with secondary education thus appear to be less conservative towards GBV across all items. It is worth mentioning that, somehow surprisingly, the younger generation does not appear to approve less GBV than the older generation — prevalence rates among men under or above 26 years old being non-significantly different to one another. Lastly, while respondents from high-FGM prevalence ethnic groups are more likely, on average, to support FGM, the difference with men from low-FGM prevalence groups is not large enough to be statistically significant.

Overall, misreporting appears homogeneously distributed in our sample. Irrespective of respondent education, age or ethnic group, support for FGM is consistently overreported when directly asked. In contrast, the underreporting of IPV justification is mostly driven by men without secondary education. Results remain robust after correcting for multiple hypotheses testing.

5.3. GBV practices

We further investigate whether the reporting of actual behavior, i.e., the perpetration of violent acts against women, suffers from response bias. Results show that the LE and DQ techniques yield similar prevalence rates: about 26% of respondents have ever committed physical violence against an adult woman, and 1% have ever raped a woman. Somewhat surprisingly, answers provided to DQs do not appear to be affected by misreporting (see Fig. 3 and Table E1 for detailed results).[20] A possible explanation for this is that male respondents perceive GBV to be generally accepted by the Malian society, making it unnecessary to conceal it. It is worth mentioning that, as discussed in Section 3.3, a possible ceiling effect may lead to an underestimation of the LE prevalence of physical violence, altering our estimation of the response bias.[21]

Turning to possible heterogeneity in prevalence rates and misreporting, Fig. 4 reports the main estimates across respondent sub-groups.[22] The LE prevalence rates are smaller for young men, for men from high FGM prevalence ethnic groups and for men educated to secondary level (respectively 22.3%, 22.7% and 22.2%) compared to older men, those from low FGM-prevalence groups, and those with lower levels of education (respectively 28%, 37.7% and 29.5%). However, none of these prevalence rates are significantly different across sub-groups (see Panel B column (3) of Table E1 in Online Appendix E). In terms of response bias, we find that men belonging to ethnic group with high FGM prevalence significantly overreport physical violence by 11.5 percentage points (see Fig. 4 and Panel B column (2) of Table E1 in Online Appendix E), but this result does not survive to correcting for multiple hypotheses testing.

[20] 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 3.3).

[21] Table D1 in Online Appendix D 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.

[22] 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 13).

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	0.443	0.161	−0.282***	0.064	−0.088
		(0.055)	(0.064)	(0.087)	[0.529]	[0.444]
	Under 26 years old	0.346	0.260	−0.086	0.020	−0.064
		(0.047)	(0.076)	(0.089)	[0.677]	[0.924]
	High FGM prevalence ethnic groups	0.362	0.307	−0.055	0.024	−0.014
		(0.084)	(0.048)	(0.099)	[1.000]	[1.000]
Child marriage	Secondary education	0.361	0.256	−0.105	−0.073	−0.053
		(0.066)	(0.075)	(0.095)	[0.529]	[0.721]
	Under 26 years old	0.313	0.323	0.010	−0.111*	0.048
		(0.057)	(0.096)	(0.105)	[0.852]	[0.172]
	High FGM prevalence ethnic groups	0.444	0.275	−0.169	0.025	−0.096
		(0.108)	(0.056)	(0.117)	[1.000]	[0.261]
FGM wife	Secondary education	0.578	0.240	−0.338***	−0.177**	−0.332***
		(0.071)	(0.085)	(0.114)	[0.058]	[0.003]
	Under 26 years old	0.476	0.323	−0.153	−0.216***	−0.325***
		(0.063)	(0.096)	(0.118)	[0.011]	[0.011]
	High FGM prevalence ethnic groups	0.403	0.439	0.036	−0.277**	−0.241***
		(0.112)	(0.055)	(0.119)	[0.064]	[0.003]
FGM daughter	Secondary education	0.632	0.398	−0.234**	−0.206***	−0.215***
		(0.064)	(0.074)	(0.099)	[0.017]	[0.017]
	Under 26 years old	0.552	0.476	−0.076	−0.211***	−0.223***
		(0.060)	(0.073)	(0.093)	[0.017]	[0.010]
	High FGM prevalence ethnic groups	0.459	0.552	0.093	−0.242**	−0.206***
		(0.104)	(0.053)	(0.115)	[0.085]	[0.003]
IPV justified	Secondary education	0.521	0.326	−0.195**	0.154**	0.039
		(0.058)	(0.076)	(0.091)	[0.048]	[0.852]
	Under 26 years old	0.424	0.464	0.040	0.093	0.124
		(0.056)	(0.089)	(0.103)	[0.362]	[0.261]
	High FGM prevalence ethnic groups	0.399	0.447	0.048	0.090	0.106*
		(0.084)	(0.057)	(0.099)	[0.529]	[0.172]

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 26 years old equals one if the respondent is 25 years old or younger; 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. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

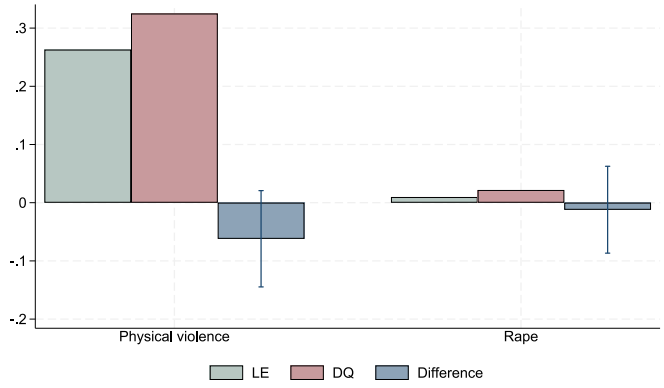

Fig. 3. GBV practices: prevalence rates and size of response biases.

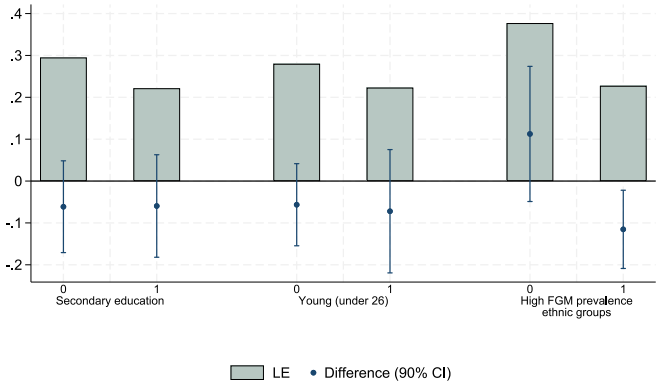

Fig. 4. Perpetration of physical violence: prevalence rates and size of response biases by sub-groups.

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.[23] The LE results show that violence is more justified (43.6%) than it is actually practiced (26.3%), suggesting that the justification of violence is not (only) an ex-post rationalization

[23] IPV justification relates to a man beating his wife if she disrespects him, while the question about the use of violence refers more generally to a man hitting a woman, regardless of whether she is his wife and of the reason for using violence.

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, 32.6% and 22.2%), and men without secondary education being those justifying it (52.1%) and practicing it the most (29.5%).

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 to avoid being victimized. Men's opinions would thus constrain women's behavior — eventually reducing the perpetration of violence. The difference in the LE rates for the perpetration of physical violence across respondent education levels would then imply that men with secondary education interact with women who "behave" — resulting in a lower IPV prevalence — while low-educated men interact with women who "misbehave" more often — resulting in a higher IPV prevalence.

Another possible interpretation for the difference between behaviors and opinions prevalence 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, the government and civil society have conducted regular campaigns against GBV in the past few years. Our finding that low-educated men are more in favor of IPV than those with higher educational attainment could signal an ongoing evolution of norms.

While other hypotheses could be drawn to explain it, we believe that the result suggesting that violence is more justified than practiced paves the way for future interesting research dedicated to understanding how social norms dynamics can shape opinions and actual behaviors differently.

6. Conclusion

Mali is one of the countries in the world where women's living conditions are the hardest amid high gender inequality. Biased gender social norms are probably among the major obstacles in achieving gender equality and empowering all women and girls. They can influence patterns of violence against women, as when individuals believe that violence is acceptable they might also directly enforce it. In this respect, opinions of men, who are most often the perpetrators of GBV, are key in the fight for gender equality, as their attitudes and behaviors may sustain the use of gender-based violence.

This article studies men's opinions and behaviors related to GBV in Mali. We conduct a LE to measure practices and support of an extensive set of GBV dimensions (wife-beating, rape, child marriage, husband's control of wife's spending and FGM) among men in Bamako. By eliciting answers to the sensitive questions using a LE technique, we limit the risk of response bias and we identify the extent to which views and behaviors are misreported with the DQ survey method.

Indirect questioning reveals that violent behaviors towards women as well as justification of GBV are widespread among men in Bamako. About one-quarter of men have already hit a woman, nearly a third is in favor of husbands exercising financial control and female child marriage. About half justify IPV, think that a good wife is a cut woman and would have their daughter's genitalia cut if they had one. Moreover, the share of respondents who support IPV is larger than the share of those engaging in its practice: the LE shows that wife-beating is more justified (43.6%) than it is practiced (26.3%).

Though violence against women is widely approved, different dimensions of GBV appear to be subject to contrasted degrees of conservatism. In particular, the control of women's economic resources as well as female child marriage are not endorsed by a majority of men. The latter result suggests that a reform raising girl's legal age of marriage at 18 could be supported by the male population. In contrast, IPV and, to a greater extent, FGM, remain popular. However, these practices being connected to each other, delaying the age of marriage could raise girls' education, possibly substituting for the need of pursuing FGM with the aim of a positive marriage match, as found by García-Hombrados and Salgado (2023), and/or possibly reducing IPV exposure in line with Wilson (2019).

Turning to response biases, our study shows that men are not reluctant to openly state whether they have already hit or raped a woman. However, we find substantial misreporting of the justification of IPV and support for FGM. Men are found to conceal justification of IPV while they overstate their support for FGM when asked directly. These contrasted patterns are consistent with respondents adapting their answers to the perceived prevalent social norm — be it more or less conservative than their real attitudes. In the case of IPV, the unfavorable perceived social norm may be related to the important efforts to raise public awareness undertaken by local NGOs and international organizations as well as by the Malian government. In the case of FGM, the persistence of a perceived conservative social norm may be related to the discourse of national Muslim organizations. 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 FGM is much lower than commonly thought, despite men preferring to conceal their true opinions. An interesting path for future research would be to 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.

CRediT authorship contribution statement

Olivia Bertelli: Project administration, Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Funding acquisition. **Thomas Calvo:** Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Funding acquisition. **Emmanuelle Lavallée:** Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Funding acquisition. **Marion Mercier:** Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Funding acquisition. **Sandrine Mesplé-Somps:** Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Funding acquisition.

Acknowledgment

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 sincerely thank the Editor and two anonymous referees for their thoughtful comments and valuable suggestions. We would also like to thank Katja Bergonzoli, Elise Huillery, Tista Mukherjee, Maëlle Stricot as well as participants in the Nanterre workshop on Gender Issues and Development, PSL WiP workshop, Dauphine workshop on the Economics of Social Norms, GATE St Etienne Economic Development Workshop in addition to participants at the 40th Journées de microéconomie appliquée, NovAfrica Conference, Oxford CSAE conference,

Table A.1
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 a cut wife* - I trust women in politics	*Is a good wife a cut wife?*
7.	- Boys should participate in household chores - *If I were to have a daughter today, I would have her cut* - 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 cut?*
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?*

Bordeaux AFSE conference, Aix ICDE conference, and seminars at IE-UFRJ and ERUDITE for their very helpful comments and suggestions. We also thank the internal review board of the Paris School of Economics for approving our experimental protocol (application # 2021 002) and the Malian Ethic committee (decision #13/2022/CE-INSP). The trial is registered at https://www.socialscienceregistry.org/trials/8275.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in the paper.

Appendix A. Lists of items

See Table A.1.

Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2025.103479.

Data availability

The data and related documentations that support the findings of this study are openly available in DataSuds repository (IRD, France) at https://doi.org/10.23708/A6GRZK. Data reuse is granted under CC-BY license.

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. Econ. Dev. Cult. Chang. 70 (4), 1565–1590.
Alcalde, M.C., Basu, S., Burrill, E., 2015. Feminist Organizing around Violence against Women in Mali, Peru, and India. Bordo, Susan and Alcalde, M Cristina and Rosenman, Ellen, University of California Press, pp. 402–412.
Alesina, A., Brioschi, B., La Ferrara, E., 2021. Violence against women: A cross-cultural analysis for Africa. Economica 88 (349), 70–104.
Amselle, J.-L., 2018. L'excision et l'homosexualité, enjeux politiques au Mali. Les Temps Mod. (2), 3–19.
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. J. Amer. Statist. Assoc. 103 (484), 1481–1495.
Asadullah, M.N., De Cao, E., Khatoon, F.Z., Siddique, Z., 2021. Measuring gender attitudes using list experiments. J. Popul. Econ. 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. Am. Political Sci. Rev. 114 (4), 1297–1315.
Blair, G., Imai, K., 2012. Statistical analysis of list experiments. Political Anal. 20 (1), 47–77.
Bulte, E., Lensink, R., 2019. Women's empowerment and domestic abuse: Experimental evidence from Vietnam. Eur. Econ. Rev. 115, 172–191.
Bursztyn, L., González, A.L., Yanagizawa-Drott, D., 2020. Misperceived social norms: Women working outside the home in Saudi Arabia. Am. Econ. Rev. 110 (10), 2997–3029.
Chuang, E., Dupas, P., Huillery, E., Seban, J., 2021. Sex, lies, and measurement: Consistency tests for indirect response survey methods. J. Dev. Econ. 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. Manag. Sci. 63 (10), 3168–3186.

Coulibaly, M., Coulibaly, M., Binto, S.F., 2022. Résumé des résultats, enquête afrobarometer round 9 Mali. Afrobarometer 91.
Cullen, C., 2022. Method matters: The underreporting of intimate partner violence. World Bank Econ. Rev. 37 (1), 49–73.
De Cao, E., Lutz, C., 2018. Sensitive survey questions: Measuring attitudes regarding female genital cutting through a list experiment. Oxf. Bull. Econ. Stat. 80 (5), 871–892.
Diabate, I., Mesple-Somps, S., 2019. Female genital mutilation and migration in Mali: Do return migrants transfer social norms? J. Popul. Econ. 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. World Bank Econ. Rev. 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. J. Interpers. 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 Opin. Q. 77 (S1), 159–172.
Holbrook, A.L., Krosnick, J.A., 2010. Social desirability bias in voter turnout reports: Tests using the item count technique. Public Opin. Q. 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. Annu. Rev. Econ. 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. 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. In: 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. World Bank Econ. Rev. 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 Glob. Heal. 3 (5), e000549.
Karlan, D.S., Zinman, J., 2012. List randomization for sensitive behavior: An application for measuring use of loan proceeds. J. Dev. Econ. 98 (1), 71–75.
Koné, O., Calvès, A.E., 2021. La mobilisation des organisations féminines en faveur du code de la famille au Mali. Autopsie d'une défaite. Cah. D'études Afr. 61 (242), 331–354.
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. Soc. Sci. Med. 266, 113326.
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. Trop. Med. Int. Health 9 (6), 737–754.
Rosenfeld, B., Imai, K., Shapiro, J.N., 2016. An empirical validation study of popular survey methodologies for sensitive questions. Am. J. Political Sci. 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. Oxf. Dev. Stud. 49 (1).
Schulz, D., 2012. Sharia and National Law in Mali. Otto, J-M., Amsterdam, pp. 529–552.
Tourangeau, R., Yan, T., 2007. Sensitive questions in surveys. Psychol. Bull. 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. J. Confl. Resolut. 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. Heal. Econ. 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 Glob. Heal. 5 (5).
Watson, C., 2014. Understanding Changing Social Norms and Practices Around Girls' Education and Marriage. Technical report, Overseas Development Institute, London.
Whitehouse, B., 2022. Left in the cold: The mirage of marriage and family law reform in post-colonial Mali. J. Leg. Anthr. 6 (1), 73–94.
Wilson, N., 2019. Socio-economic status, demographic characteristics and intimate partner violence. J. Int. Dev. 31 (7), 632–657.