NUMBER 6

CMI WORKING PAPER

DECEMBER 2025

Photo by Yohannes Minas (on Unsplash)

AUTHORS

Matthew Gichohi
CMI

Andreas Kotsadam
Ragnar Frisch Centre for Economic Research

Mette Løvgren
Oslo Metropolitan University

Charlotte Ringdal
CMI

Espen Villanger
CMI

Does it matter what others think? Information, norms, and female genital mutilation in Ethiopia

CMI CHR. MICHELSEN INSTITUTE

Does it Matter What Others Think? Information, Norms, and Female Genital Mutilation in Ethiopia*

Matthew Gichohi[1], Andreas Kotsadam[2], Mette Løvgren[3], Charlotte Ringdal[1], and Espen Villanger[1]

[1]CMI
[2]Ragnar Frisch Centre for Economic Research
[3]Oslo Metropolitan University

Abstract

Using data on over 3,000 mothers and fathers in five Ethiopian regions, we study how misperceived social norms about female genital mutilation (FGM) relate to cutting decisions and whether norm information changes them. At baseline, many parents who say FGM should stop have a cut daughter and still plan to cut, parents overestimate local support for FGM, and these misperceptions are correlated with intended cutting. Motivated by this, we field an information experiment that randomly informs parents about opposition to FGM in a nearby community. The intervention has no detectable effects on reported beliefs about others' attitudes or on respondents' own attitudes. For fathers, however, it modestly reduces stated intentions to circumcise: plans to cut a daughter fall by 8 percentage points and willingness to cut a hypothetical daughter by 4 percentage points, though these effects are only partly robust to multiple-testing adjustments. We find no effects for mothers.

*We thank Selim Gulesci for sharing their survey and seminar participants at Oxford University, Kent University, 2nd Welfare and Policy Conference, Bordeaux School of Economics, and FAIR at NHH for valuable comments. A pre-analysis plan was archived at the registry for randomized controlled trials in economics held by The American Economic Association before any midline data was collected. The plan is available here: https://andreaskotsadam.wordpress.com/wp-content/uploads/2024/11/pap_fgm_info.pdf. Funding from the Research Council of Norway (project 324885) and Norad is acknowledged. The research presented in this paper has IRB approval from the ethical committee at CMI and was reviewed and approved by the Norwegian Centre for Research Data (ref number 159019). Authorship is listed alphabetically. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

1 Introduction

Female genital mutilation (FGM) is a severe violation of girls' bodily integrity that affects hundreds of millions of women worldwide and continues to put millions of girls at risk every year (UNICEF, 2024). Despite decades of advocacy, legal bans, and programmatic efforts, progress has been uneven and slow (United Nations Population Fund (UNFPA), 2025). In Ethiopia, for example, around two-thirds of women aged 15–49 have undergone FGM, and new cohorts of girls continue to be cut (Alemu, 2021). Against this backdrop, understanding why parents persist in a practice that carries clear health risks and is increasingly contested, both internationally and within many communities, is crucial for designing effective policies.

A prominent explanation is that FGM is sustained by social norms and expectations (Efferson et al., 2015). Parents may fear that an uncut daughter will face stigma, lower marriage prospects, or social sanctions, even if they personally oppose the practice (UNICEF Innocenti Research Centre, 2005; Shell-Duncan et al., 2011). In such a setting, the key constraint is not necessarily parents' own preferences, but what they think others in their community believe and do. A growing literature on misperceived norms shows that people often overestimate conservative or harmful attitudes among peers, and that correcting these misperceptions can shift behavior (Bursztyn, González and Yanagizawa-Drott, 2020; Haaland, Roth and Wohlfart, 2023). Yet we know little about whether similar mechanisms operate for FGM, and almost nothing about whether simple information interventions can change parents' FGM decisions.

We address these questions using new data on 1,635 mothers and 1,510 fathers from 78 kebeles in five Ethiopian regions (Somali, Afar, Sidama, Amhara, and Oromia). Our dataset links mothers and fathers within the same household and records, for each of them, their own stance on FGM, their perceptions of what other men and women in the community believe, the FGM status of each of their daughters, and their intentions regarding both

current and hypothetical future daughters. This unusually rich combination of attitudes, beliefs, and actual or intended child outcomes enables us to examine how parents jointly navigate decisions around a harmful cultural practice. In the baseline, we documented three descriptive facts that motivate our experimental intervention. First, there is a substantial gap between stated attitudes and behavior: many parents who oppose FGM nevertheless have at least one cut daughter, and a sizeable share of parents of uncut daughters still plan to cut them in the future. Second, parents substantially overestimate local support for FGM. When asked what share of women or men in their village think that a girl should be circumcised, most respondents report much higher support than we observe in the same communities. Third, these misperceptions are strongly correlated with intentions to cut: parents who overestimate others' support for FGM are much more likely to say that they plan to circumcise a daughter.

These patterns suggest that FGM may persist not only because of true social pressure, but also because of *incorrect beliefs* about the strength of that pressure. Motivated by this insight, we designed a pre-registered information provision experiment. Approximately 2.5 years after the baseline survey, we re-interviewed the original respondents and randomly informed a subset of them about actual opposition to FGM in a "nearby community" in their region. The information was based on our own baseline data: respondents were told that in a neighboring kebele, X out of 100 women (or men) disagreed with the statement "*A girl in our village should be circumcised,*" where X ranged from 57 to 97. To avoid inadvertently encouraging FGM, we only included parents for whom this information was more progressive than their prior beliefs, and we excluded the most progressive communities where opposition was already very high.

The intervention was deliberately light-touch: a short script read by survey enumerators in the middle of the follow-up interview. We focus on four pre-specified outcomes measured at follow-up: (i) perceived community support for FGM, (ii) own attitude towards cutting,

(iii) plans to circumcise any uncut daughter under 12, and (iv) the intention to circumcise a hypothetical future daughter (a vignette). Using an individual-level randomized design with region fixed effects and optimally selected baseline controls, we estimate the causal effect of receiving information about others' opposition to FGM on these outcomes.

Three main results emerge. First, the information treatment does not change respondents' reported beliefs about how many men and women in their community support FGM, nor does it change their own expressed attitudes. The null results are precise: using pre-registered equivalence tests, we can rule out even moderately sized changes in average attitudes and beliefs. Second, despite the absence of detectable belief updating, the information has economically meaningful effects on men's reported intentions. Among fathers, the treatment reduces the probability of planning to circumcise at least one daughter by 8.1 percentage points (about 22 percent of the control mean) and lowers the probability of circumcising a hypothetical future daughter by 4.2 percentage points (around 14 percent). For mothers, in contrast, we find no evidence of changes in intentions. Third, among fathers, both messages, whether they provided information about women's or men's opposition to FGM in nearby communities, produce negative effects on intended cutting, although only one is statistically significant. We cannot statistically distinguish the two treatment effects from each other.

Taken together, our descriptive and experimental findings speak to several literatures. First, we contribute to the literature on FGM and social norms. The dominant "social coordination" view argues that FGM is sustained by an equilibrium in which families coordinate on cutting girls until a critical mass of families abandons the practice (Mackie, 1996; Shell-Duncan et al., 2011). Recent work, however, questions whether FGM in fact behaves like a classic coordination norm (Efferson et al., 2015; Novak, 2020; Congdon Fors, Isaksson and Lindskog, 2021; Fors, Isaksson and Lindskog, 2024; Gulesci et al., 2021). Our baseline results show that misperceived norms are closely linked to parents' cutting decisions even when

many parents privately oppose FGM, and the experimental results indicate that updating information about others' opposition can reduce intended cutting among men, without necessarily shifting their stated attitudes. This highlights the role of beliefs about others, not only actual sanctions, in sustaining harmful practices.

Second, we add to the evidence on interventions aimed at changing harmful household behaviors. Existing initiatives, from community empowerment programs such as Tostan in Senegal (Camilotti, 2016) to broader religious and NGO campaigns (Fors, Isaksson and Lindskog, 2024), tend to be intensive, long-running, and costly. Most closely related to our study, Ferreira et al. (2024) conduct a large field experiment in Somalia that combines belief-elicitation, revelation of community support for abandoning infibulation, and a public-declaration coordination device; they find that correcting misperceptions substantially reduces the most harmful form of cutting, albeit with some substitution toward a milder type. Our intervention is deliberately more light-touch and is implemented at the individual rather than meeting level, and we focus on how such norm-based information differentially affects mothers and fathers within the same household.

Third, we contribute to the economics-of-the-household literature by using matched data on mothers and fathers to study how parents in the same household form beliefs and state intended behavior. While previous work stresses women's central role in organizing cutting (e.g. Ferreira et al., 2024), we find that in our Ethiopian sample only fathers adjust their stated intentions in response to norm-based information, even though mothers' attitudes and beliefs are at least as progressive. This illustrates that interventions that work through perceived social norms can have gender-differentiated effects within households.

The rest of the paper proceeds as follows. Section 2 describes the context and the data, including the three motivating observations on attitudes, misperceptions, and behavior. Section 3 details the experimental design and implementation, including how we construct the information treatment, and the empirical strategy. Section 4 reports the main experimen-

tal results and heterogeneity analysis. Section 5 discusses mechanisms and the substantial changes in FGM-related attitudes over time, and Section 6 concludes with policy implications.

2 Context and data

In this section, we briefly describe the broader context of FGM in Ethiopia and in the five regions where our study takes place, before turning to the baseline survey and the main patterns that motivate our experimental intervention.

2.1 FGM in Ethiopia and study regions

FGM remains widespread in Ethiopia. Using the three most recent Demographic and Health Surveys, Alemu (2021) documents that around 65% of Ethiopian women aged 15–49 have undergone some form of FGM, with only modest declines over time. In the five regions where our study is implemented, Somali, Afar, Oromia, Amhara, and Sidama, the share of women aged 15–49 who are cut is 99%, 91%, 76%, 62%, and 62%, respectively (Central Statistical Agency (CSA) [Ethiopia] and ICF, 2016). These regions therefore represent some of the highest-prevalence areas in the country and are central to understanding the persistence of the practice.

The form of cutting also varies substantially across regions. According to the 2016 DHS, infibulation (WHO type III) is highly prevalent in Somali and Afar, where 73% and 64% of cut women, respectively, report having undergone this most severe form. In contrast, infibulation is rare in Oromia, Amhara, and Sidama, where it affects fewer than 10% of cut women. In these regions, cutting more commonly involves removal of genital flesh without sewing (WHO types I and II) (Central Statistical Agency (CSA) [Ethiopia] and ICF, 2016). This pattern is consistent with broader evidence that the most severe forms of FGM are geographically

concentrated, while less extensive forms are more widespread (UNICEF, 2024).

Age at cutting shows another dimension of regional heterogeneity. In Afar and Amhara, FGM is typically performed very early in life: 90% of cut women in Afar and 92% in Amhara report having been cut before the age of five. In Somali, the practice tends to occur later in childhood, with 61% of cut women reporting that they were cut between ages five and nine. In Sidama and Oromia, the timing is more evenly spread across childhood and early adolescence, with no single age group dominating (Central Statistical Agency (CSA) [Ethiopia] and ICF, 2016). These differences in timing are important for interpreting both existing prevalence levels and the scope for interventions that target parents' intentions for younger or not-yet-born daughters.

Economic mechanisms have been proposed as factors sustaining FGM, such as the role of traditional "cutters" who depend on the practice for their livelihood and the belief that cutting increases a girl's marriage prospects. However, empirical evidence from Ethiopia challenges these notions. For instance, Gibson et al. (2023) find no significant association between being cut and receiving a higher bride price in rural Ethiopian communities. Further, from our own qualitative work, it does not look like cutters have an economic incentives to keep the practice.

Ethiopia has adopted a range of policies and programs to combat FGM. The Revised Criminal Code of 2004 explicitly criminalizes the performance and procurement of FGM under Articles 565–570, and infibulation is singled out for particular penalties. In addition, the Ministry of Health issued a directive in 2017 banning the medicalization of FGM in health facilities. At the program level, Ethiopia has been a priority country in the UNFPA–UNICEF Joint Programme on the Elimination of FGM, launched in 2008, which supports community education, engagement with religious and traditional leaders, and strengthening of national response systems (UNFPA and UNICEF, 2024). Available survey data indicate a gradual decline in FGM prevalence among women 15–49 over the past two decades, and a much

sharper decline among their daughters, although levels remain very high in several regions (e.g. Central Statistical Agency (CSA) [Ethiopia] and ICF, 2016; Alemu, 2021; UNICEF, 2024). Enforcement of the law has been uneven, and most existing interventions have not explicitly targeted misperceived local norms, which is the focus of the present study.

2.2 Baseline survey and sampling

We conducted a baseline household survey in collaboration with Norwegian Church Aid and Save the Children. The survey covered 3,145 parents, 1,635 mothers and 1,510 fathers, in 78 kebeles across five Ethiopian regions: Somali, Afar, Sidama, Amhara, and Oromia. The largest share of respondents live in Somali (44%), while Amhara and Oromia contribute around 9–10% each.

Within selected kebeles, households were eligible for inclusion if they had at least one child below the age of 13. We implemented a random walk within each kebeles to recruit households for the study. In each eligible household, we aimed to interview both a woman and her spouse; when one partner was absent, we interviewed only the available parent. This design yields a sample in which we observe both parents for most households.

Interviews were conducted face-to-face by an independent professional survey team experienced in collecting data on sensitive topics such as FGM. Enumerators followed a structured questionnaire that included detailed modules on household composition, the FGM status of all daughters, parents' own attitudes towards FGM, and their beliefs about other community members' attitudes and behavior. We also included a social desirability scale and a list experiment module, which we later use to assess potential reporting bias.

2.3 Baseline descriptives and key patterns

Table 1 summarizes the main characteristics of the 3,145 respondents interviewed at baseline (1,635 mothers and 1,510 fathers). On average, respondents are 39 years old and have 3.8 children living in the household, of whom around 2.1 are girls. Fathers are older, more likely to be literate and employed, and have slightly more years of schooling than mothers. At the same time, attitudes and beliefs related to FGM are remarkably similar across genders: around 41% of respondents agree that "a girl in our village should be circumcised", and roughly half state that they would circumcise a hypothetical future daughter (vignette measure). Respondents also perceive FGM as common in their community (mean response close to "common" on the 1–3 scale), and on average believe that about half of men and women in their village support cutting. Finally, about 60% of parents report having at least one circumcised daughter, with an average of one cut daughter per household.

Table 1: Descriptives

	(1) All		(2) Men		(3) Women	
	Mean	SD	Mean	SD	Mean	SD
Age	39.01	(9.19)	42.96	(9.51)	35.36	(7.17)***
Literate	0.41	(0.49)	0.57	(0.50)	0.25	(0.44)***
Employed	0.79	(0.41)	0.95	(0.23)	0.64	(0.48)***
Education	2.45	(2.62)	2.95	(3.12)	2.00	(1.93)***
Number of children	3.80	(1.67)	3.83	(1.68)	3.78	(1.67)
Number of girls	2.07	(0.98)	2.08	(0.98)	2.06	(0.98)
Vignette: Hypothetical daughter would be cut	0.50	(0.50)	0.46	(0.50)	0.53	(0.50)***
Own attitude: Agree that village girls should be cut	0.41	(0.49)	0.41	(0.49)	0.42	(0.49)
How common is FGM (1-3)	1.90	(0.80)	1.88	(0.80)	1.91	(0.80)
Belief: Percent women that agree	53.00	(34.21)	52.80	(34.02)	53.18	(34.39)
Belief: Percent men that agree	50.45	(33.47)	49.52	(33.71)	51.32	(33.25)
Any cut daughter	0.60	(0.49)	0.60	(0.49)	0.61	(0.49)
Number of cut daughters	1.02	(1.07)	1.02	(1.08)	1.01	(1.06)
N	3145		1510		1635	

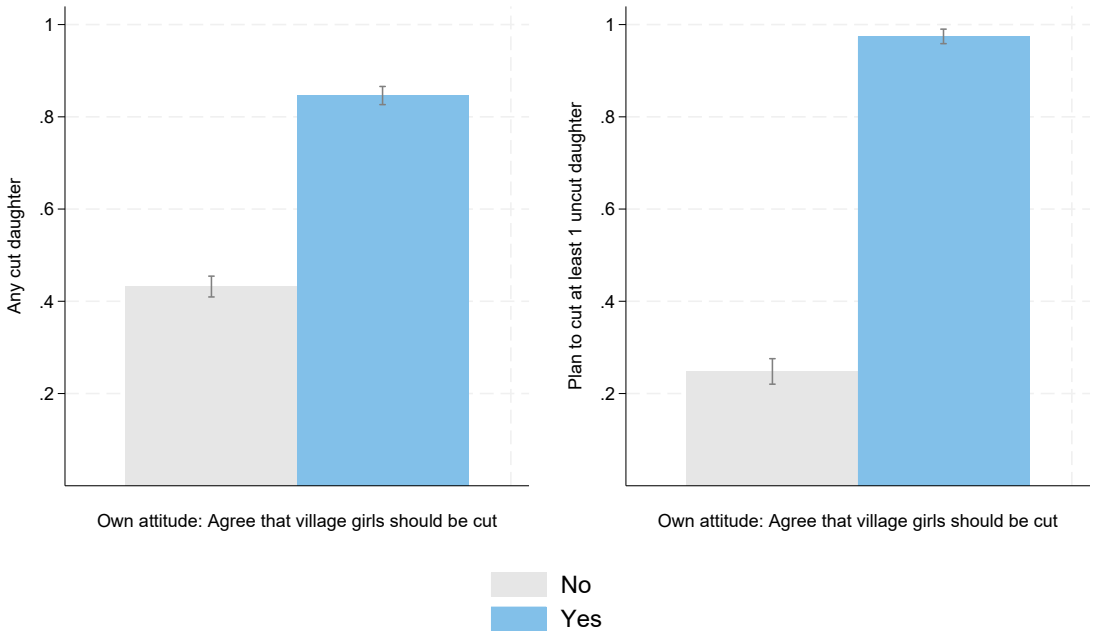
Notes: The sample consists of everyone at baseline. Tests of differences for each variables are indicated by by $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Beyond these summary statistics, the baseline data reveal three empirical regularities that are central for the design of the information intervention.

Observation 1: Many parents cut or plan to cut their daughters even if they oppose FGM. Figure 1 plots current and planned FGM status of daughters by whether parents agree that girls in their village should be circumcised. Among parents who support FGM, more than 80% have at least one daughter who is already cut, and almost all parents with uncut daughters intend to circumcise at least one of them in the future. Strikingly, even among parents

who do not agree that girls should be circumcised, almost 50% already have a cut daughter, and about a quarter of parents with uncut daughters still plan to circumcise at least one of them. This gap between stated opposition and behavior suggests that individual disapproval is often not sufficient to prevent cutting.

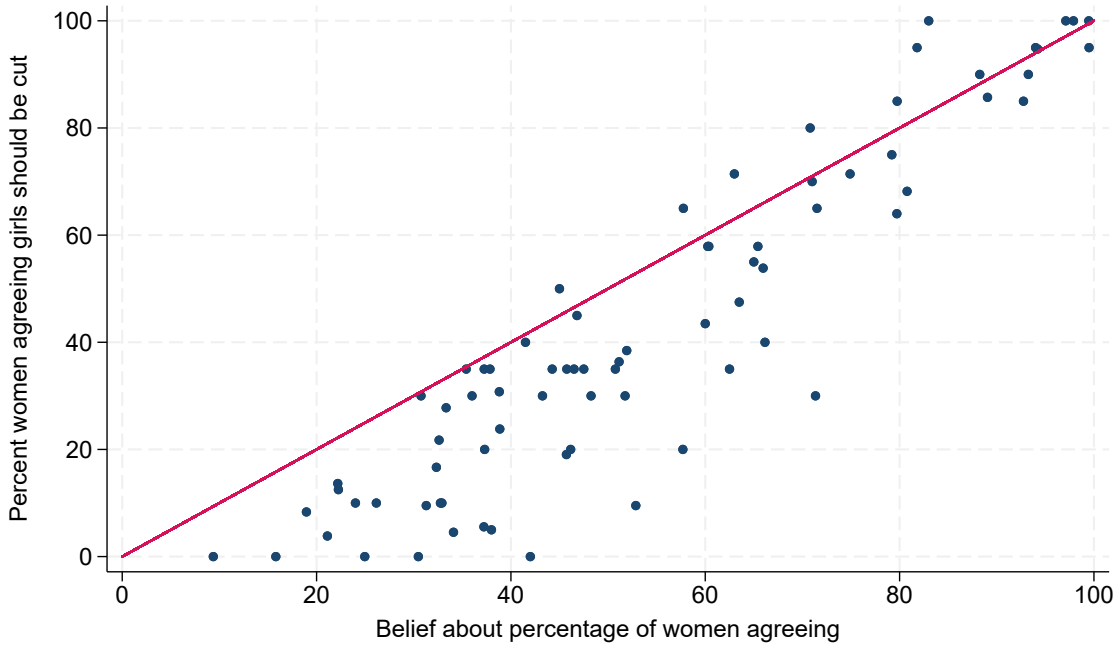
Figure 1: Observation 1: Many parents cut their daughters even if they oppose the practice



Notes: Full sample at baseline for cutting, only those with at least one uncut daughter for plans. 95 percent confidence intervals.

Observation 2: Parents tend to overestimate community support for FGM. Figure 2 compares, at the kebele level, the actual share of women who agree that girls should be circumcised with the average belief about that share. Most points lie well below the 45-degree line, indicating that respondents systematically overestimate how many women in their community support FGM. A similar pattern holds when we use men's attitudes and beliefs (Appendix Figure A.1).

Figure 2: Observation 2: Parents tend to overestimate the support of FGM in their community

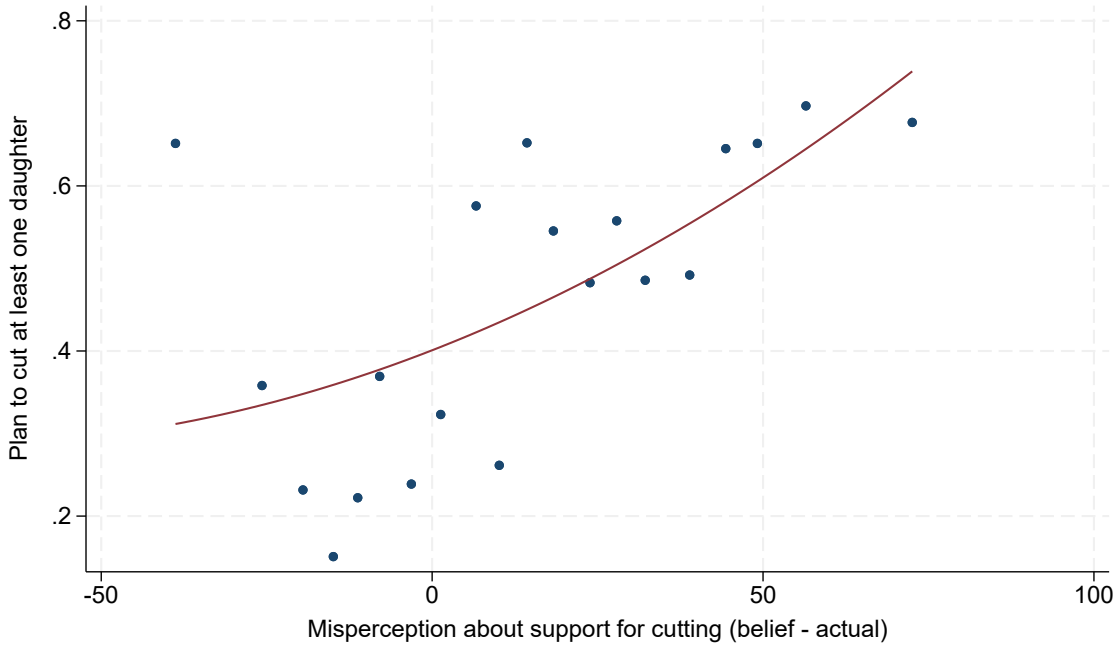


Notes: Kebele level averages. Full sample of women at baseline. 45-degree line in red.

Observation 3: Misperceptions are strongly correlated with intentions to cut daughters. Finally, Figure 3 documents a strong positive association between the extent of misperception, defined as the difference between an individual's belief about women's support for cutting and the actual share of women who support it in the kebele, and the probability that the respondent plans to circumcise at least one daughter in the future. Parents who substantially overestimate community support are much more likely to intend to cut.

Taken together, these three observations suggest that high FGM prevalence in our study areas is not only driven by individual preferences, but also by misperceived social norms. This motivates the experimental design described in Section 3.

Figure 3: Observation 3: There is a strong correlation between misperceptions and intention to cut a daughter in the future



Notes: Full sample at baseline. The misperceptions are calculated as beliefs about attitudes of women minus actual attitudes, for women the actual attitudes removes the respondent herself.

2.4 Measures of FGM and reporting issues

Self-reported attitudes and behaviors may suffer from a social desirability bias if people are ashamed of their behavior or if they feel that the enumerators may think less of them if they have certain attitudes. Such worries are largest with sensitive topics. In Appendix Section A.2, we show that reporting issues do not seem to be a great concern in our data. To summarize what we find there, we believe that the degree of underreporting in our data is likely limited due to the careful data collection. One indication of this is the high actual reported prevalence and the high acceptance of FGM in the data. While we find that individuals scoring high on a social desirability index report differently, this may be due to other factors linked to both prevalence and social desirability. We find the within

household reporting results, where parents simultaneously but separately report very similar values, to be the strongest evidence. In addition, we conduct a list experiment which does not indicate reporting bias, even for the respondents with a high social desirability. Our results are consistent with other findings showing misreporting to not be a concern in FGM studies. For instance, Elmusharaf, Elhadi and Almroth (2006) compare self-reports with clinical examinations in Sudan and find no misreporting in the levels of FGM reported.

3 Experimental design and empirical strategy

3.1 Follow-up survey and experimental sample

We conducted a follow-up survey in June–July 2023, approximately 2.5 years after the baseline survey. Thanks to a careful tracking protocol, attrition was very low: we were able to re-interview 1,622 of the 1,635 mothers in the baseline sample (0.8% attrition) and 1,471 of the 1,510 fathers (2.6% attrition).[1]

The experimental sample is a subset of the follow-up respondents. To construct it, we first calculate for each kebele the baseline share of women and men who *disagree* with the statement "A girl in our village should be circumcised." Within each region, we then identify "a nearby community": the kebele with the highest disapproval rate that is less than 100%. Let X denote the corresponding share of women (or men) who disapprove in that nearby kebele; in our data, X ranges from 57 to 97.

To avoid providing information that could inadvertently make respondents more favorable to FGM, we restrict attention to parents for whom the nearby-community statistic is (weakly) more progressive than both their own beliefs and their own kebele. Specifically, we drop (i)

[1] In the pre-analysis plan, we anticipated higher attrition and proposed to replace missing households with similar households from the same kebeles. Given the very low attrition rates, we did not implement this replacement procedure.

all kebeles whose own disapproval rate exceeds X in their region, and (ii) individuals who, at baseline, already believed that more than X percent of women (or men) in their village disapprove of FGM. These restrictions primarily exclude the most progressive communities and the most progressive individuals within remaining communities (see Figures A.2 and A.3), reducing external validity but ensuring that our progressive information treatment cannot move anyone "in the wrong direction." After applying these criteria, the experimental sample consists of 1,251 mothers and 1,162 fathers.

All randomization and treatment assignment described below are carried out within this experimental sample.

3.2 Information treatment

The intervention is a light-touch information treatment embedded in the follow-up survey. It is motivated by the baseline evidence that many parents substantially overestimate support for FGM in their community and that these misperceptions are strongly correlated with intended cutting of daughters.

For respondents assigned to receive information, enumerators read a short script providing accurate information about opposition to FGM in a nearby community in the same region. The script is:

> Last time, we visited many people in addition to you in several regions of Ethiopia. All of them had children below the age of 13. One of the questions we asked them was whether they agree or disagree with the following statement: *"A girl in our village should be circumcised."* In one of the communities close to yours, X out of 100 women/men said that they disagreed. This means that most of the women/men we interviewed in that community do not think that a girl should be circumcised.

The value of X is taken from our own baseline data, as described above. To maintain credibility, we never set $X = 100$ even when disapproval in the nearby kebele is complete; instead we use the highest observed disapproval rate below 100% in that region.[2]

Among treated respondents, we randomly vary the reference group: for some, X refers to the share of *women* in the nearby community who disagree with cutting; for others, it refers to the share of *men*. This allows us to test whether information about women's versus men's opposition has different effects on mothers and fathers. In the Afar region, however, the share of men supporting FGM at baseline was above 50% in all kebeles, so providing information about men would have implied a local majority in favor of cutting. To avoid this, Afar respondents who receive information are only told about women's disapproval.

3.3 Randomization and balance

Within each region and separately by respondent gender, we randomly assign individuals in the experimental sample to one of two conditions:

1. Information treatment (2/3 of sample): respondents receive the script above, with X referring either to women's or men's disapproval in a nearby community.

2. Control group (1/3 of sample): respondents receive no information; the interviewer proceeds directly to the next survey module.

Randomization was implemented at the individual level using random numbers pre-generated by the research team and incorporated into the survey software. In all regions except Afar, the information treatment is further split roughly in half between the "women information" and "men information" versions; in Afar all treated respondents receive the "women information" version.

[2]We were concerned that respondents might find it implausible that literally nobody in a community supported FGM.

In the empirical analysis, we will primarily compare all respondents assigned to any information treatment to the control group, and then explore heterogeneity by respondent gender and by whether the information refers to women or men in the nearby community.

To test for balance, we regress treatment on the independent variables from the baseline, both individually and together, while controlling for the strata variables (region fixed effects). Continuous variables are not dummy-coded in the balance tests. Table 2 shows that there are small and statistically insignificant differences between treatment and control groups. The F-test of whether the control variables jointly predict treatment status, displayed at the bottom of Table 2, shows that the randomization worked.

Table 2: Balance

	(1) Treated		(2) Control	
	Mean	SD	Mean	SD
Control variables				
Age	38.69	(8.78)	38.62	(9.29)
Man	0.48	(0.50)	0.46	(0.50)
Literate	0.38	(0.48)	0.36	(0.48)
Employed	0.76	(0.43)	0.79	(0.41)
Education	2.18	(2.42)	2.14	(2.26)
Number of children	3.81	(1.71)	3.83	(1.70)
Number of girls	2.04	(0.97)	2.06	(0.98)
Vignette: Hypothetical daughter would be cut	0.59	(0.49)	0.60	(0.49)
Own attitude: Agree that village girls should be cut	0.52	(0.50)	0.52	(0.50)
How common is FGM	2.08	(0.78)	2.06	(0.78)
Religion supports FGM	0.30	(0.46)	0.30	(0.46)
Belief: Percent women that agree	62.38	(30.88)	62.46	(30.74)
Belief: Percent men that agree	59.60	(30.21)	59.97	(30.08)
N	1555		789	

Notes: All control variables are measured before the experiment. Tests of differences for each variables with controls for region fixed effects are indicated by $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$. An F-test of the predictability of treatment of all control variable coefficients jointly, after controlling for region fixed effects, yields a p-value of 0.40.

3.4 Empirical Strategy

To estimate individual-level treatment effects, we use an intention-to-treat (ITT) specification of the form

$$Y_{2i} = \alpha + \beta \, \text{Info}_i + \delta' X_{1i} + \lambda_r + \varepsilon_i, \tag{1}$$

where Y_{2i} denotes the outcome for individual i measured at follow-up ($t = 2$), Info_i is an indicator that the respondent was randomly assigned to receive any information treatment, X_{1i} is a vector of baseline covariates ($t = 1$), and λ_r are region fixed effects. In some specifications, we replace Info_i with two indicators for the "women information" and "men information" versions of the treatment.

The baseline covariates X_{1i} always include the baseline value of the outcome when available, along with additional controls selected from the full set of variables in section A.3. Following our pre-analysis plan, we use the double-selection LASSO procedure of Belloni, Chernozhukov and Hansen (2014) to choose this set of controls in a data-driven way. This approach is designed to increase precision, especially in the presence of rich baseline information, while maintaining valid inference for β.

Region fixed effects λ_r are always included and are not subject to variable selection. All continuous baseline variables considered as candidates for X_{1i} are discretized into indicator variables so that each category contains at least five percent of the sample. When a baseline covariate has missing values, we set the missing entries to zero and add a separate indicator for missingness, so that observations with missing data are not dropped.

We estimate equation (1) using ordinary least squares and report robust standard errors. All results are ITT estimates with respect to the random assignment of the information treatment.

4 Results

In this section, we present the intention-to-treat effects of the information intervention on beliefs, attitudes, and intended cutting. We begin with the main effects for the full experimental sample and by respondent gender, and then turn to heterogeneity by information type (information about women versus men in nearby communities) and to the role of social

desirability in shaping responses.

4.1 Main results

Table 3 reports the effects of the information treatment on the four primary pre-specified outcomes. Panel (A) pools men and women. Columns (1) and (2) show that providing information about opposition to FGM in a nearby community has essentially no impact on beliefs about others' attitudes or on respondents' own attitudes: point estimates are close to zero and statistically insignificant. In line with our pre-analysis plan, we implement equivalence tests using two one-sided t-tests (TOST) and report the corresponding 90 percent confidence intervals in the table. These intervals show that the data allow us to rule out relatively large positive or negative effects on beliefs and own attitudes.

For intended behavior, the point estimates are larger in magnitude. In the pooled sample, the treatment reduces the likelihood that respondents plan to cut any of their daughters by about 4 percentage points (Column (3)) and reduces the probability that they would cut a hypothetical future daughter in the vignette question by about 2 percentage points (Column (4)), corresponding to an 8 percent decline relative to the control mean. The effect on the vignette is statistically significant at the 10 percent level, whereas the estimate for plans to cut current daughters is not. However, when we correct for testing four primary outcomes using the stepwise multiple-hypothesis procedure in Fink, McConnell and Vollmer (2014), none of the pooled estimates remains statistically significant.

Panels (B) and (C) separate the sample by respondent gender. For men, we again find no effects on beliefs or own attitudes, but sizeable changes in intended cutting. Relative to the control group, treated men are 8.1 percentage points less likely to report that they plan to cut a daughter (a reduction of about 22 percent) and 4 percentage points less likely to state that they would cut a hypothetical daughter in the vignette. For women, in contrast, all four outcomes show small and statistically insignificant effects, and the confidence intervals

allow us to rule out moderately sized changes in beliefs and attitudes. Thus, the information treatment appears to shift intended behavior for men only.

We again account for multiple testing of the four primary outcomes. Following Fink, McConnell and Vollmer (2014), the effect on men's vignette responses remains statistically significant at the 5 percent level when we consider the four outcomes as a family, whereas the effect on men's plans to cut their daughters is significant only at the 10 percent level.[3] Given that the information treatment has no detectable impact on beliefs or own attitudes, this pattern suggests that men become less willing to state that they will cut their daughters, but we do not find evidence that they update their stated views about social norms around FGM.

4.2 Information type

The experiment also allows us to examine whether it matters whose opinions are revealed. In most regions, treated respondents were randomly assigned to receive information either about women or about men in a nearby kebele.

Table 4 displays the results for male respondents. Across both outcomes for men; plans to cut a daughter and the vignette. the point estimates for the "information about women" and "information about men" arms are qualitatively similar and always negative. In some specifications, one of the two coefficients is statistically significant while the other is not; however, the estimates are not statistically distinguishable from each other. This means that we cannot conclude that learning about women's attitudes has a different effect than learning about men's attitudes, even though one arm may appear more precisely estimated in a given regression.

[3]If we instead treat the eight coefficients for men and women together as belonging to the same family of hypotheses, only the vignette effect for men remains statistically significant, and only at the 10 percent level.

Table 3: Main results.

(A) All

	(1) Belief	(2) Own attitude	(3) Plan	(4) Vignette
Information treatment	-0.73	-0.0060	-0.040	-0.021*
	(0.76)	(0.0047)	(0.028)	(0.011)
P-value	(0.335)	(0.203)	(0.158)	(0.070)
90% confidence interval	[-1.97,0.51]	[-0.014,0.0017]	[-0.086,0.0066]	[-0.039,-0.0019]
Control mean	35.50	0.14	0.33	0.25
N	2344	2344	520	2344
Controls	Lasso	Lasso	Lasso	Lasso
Sample	All	All	All	All

(B) Men

	(1) Belief	(2) Own attitude	(3) Plan	(4) Vignette
Information treatment	-1.62	-0.0069	-0.081**	-0.042**
	(1.07)	(0.0074)	(0.040)	(0.017)
P-value	(0.130)	(0.354)	(0.044)	(0.012)
90% confidence interval	[-3.38,0.14]	[-0.019,0.0053]	[-0.15,-0.015]	[-0.069,-0.014]
Control mean	36.77	0.17	0.36	0.28
N	1102	1102	246	1102
Controls	Lasso	Lasso	Lasso	Lasso
Sample	Men	Men	Men	Men

(C) Women

	(1) Belief	(2) Own attitude	(3) Plan	(4) Vignette
Information treatment	0.095	-0.0072	0.016	-0.0051
	(0.016)	(0.0057)	(1.07)	(0.039)
P-value	(0.929)	(0.203)	(0.688)	(0.742)
90% confidence interval	[-1.66,1.85]	[-0.017,0.0021]	[-0.048,0.080]	[-0.031,0.021]
Control mean	34.43	0.12	0.31	0.23
N	1242	1242	274	1242
Controls	Lasso	Lasso	Lasso	Lasso
Sample	Women	Women	Women	Women

Notes: The sample consists of everyone in the experimental sample in panel (A), all men in Panel B, and all women in Panel (C). Robust standard errors in parentheses. P-values are also in parentheses and indicated by $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$. 90% confidence intervals in brackets.

Table 4: Detailed treatment, men

	(1)	(2)	(3)	(4)
	Belief	Own attitude	Plan	Vignette
Information about men	-2.11	-0.0017	-0.088*	-0.037
	(1.39)	(0.0098)	(0.047)	(0.024)
Information about women	-1.33	-0.010	-0.075	-0.044**
	(1.18)	(0.0077)	(0.046)	(0.017)
Control mean	36.77	0.17	0.36	0.28
N	1102	1102	246	1102
Controls	Lasso	Lasso	Lasso	Lasso
Sample	Men	Men	Men	Men

Notes: Robust standard errors in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

For women (see Table 5), we do not find systematic differences by information type. Point estimates for both information arms are small and not statistically different from zero for all four primary outcomes, and we can reject large differences between learning about men versus women.

Table 5: Detailed treatment, Women

	(1)	(2)	(3)	(4)
	Belief	Own attitude	Plan	Vignette
Information about men	-0.71	-0.0021	0.036	-0.020
	(1.38)	(0.0077)	(0.044)	(0.021)
Information about women	0.59	-0.010*	-0.0039	0.0039
	(1.17)	(0.0058)	(0.047)	(0.017)
Control mean	34.43	0.12	0.31	0.23
N	1242	1242	274	1242
Controls	Lasso	Lasso	Lasso	Lasso
Sample	Women	Women	Women	Women

Notes: Robust standard errors in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Taken together, the evidence suggests that the content of the information, whether it refers to men or women in the nearby community, does not have significantly different effects. The reductions in intended cutting we observe for men appear broadly similar across the two information types, although sampling variation means that only one of the two estimates is statistically significant in any given specification.

4.3 Social desirability

As FGM is a highly sensitive topic, an important concern is that our survey measures may be affected by social desirability bias. We measure social desirability at *baseline* using the shortened Marlowe-Crowne social desirability scale (Crowne and Marlowe, 1960; Reynolds, 1982), based on 13 "saintlike" traits such as always being a good listener or always admitting one's mistakes. Following Dhar, Jain and Jayachandran (2018), we define an indicator for *high social desirability* equal to one for respondents who endorse 8 or more of the 13 traits (See Section A.2 for more details).

We use this baseline indicator, which is pre-determined with respect to the intervention, to examine whether social desirability is related to the level of reported attitudes and intentions, and whether it moderates the effect of the information treatment. Tables 6 and 7 report results from augmenting our main specification with the high-social-desirability indicator and its interaction with the information treatment, separately for men and women.

For men, high baseline social desirability is clearly associated with more progressive reported intentions. Relative to men with lower social desirability, those classified as high social desirability are about 15 percentage points less likely to report that they plan to cut a daughter and around 6 percentage points less likely to state that they would cut a hypothetical daughter (Columns (3)–(4) in Table 6). A similar pattern holds for women: high-social-desirability women are roughly 7 percentage points less likely to say they would cut a hypothetical daughter (Table 7, Column (4)). These correlations are exactly what one would expect if individuals who tend to give socially approved answers are also more likely to report that they oppose FGM and will not cut their daughters.

Crucially, however, we find little evidence that social desirability systematically moderates the impact of the information treatment. For men, the interaction terms between the treatment and high social desirability are small and statistically insignificant for beliefs, own attitudes, plans, and vignette intentions. For women, the only significant interaction appears for own attitudes (Column (2) of Table 7): the treatment slightly reduces stated support for FGM among high-social-desirability women by about 2.5 percentage points, but the magnitude is small. For the behavioral outcomes (plans to cut and vignette), the interaction terms are again small and statistically insignificant.

Table 6: Social desirability. Men

	(1)	(2)	(3)	(4)
	Belief	Own attitude	Plan	Vignette
Information treatment	-2.55	-0.0019	-0.15*	-0.033
	(1.60)	(0.013)	(0.085)	(0.030)
High social desirability	-1.93	-0.0029	-0.15*	-0.062**
	(1.78)	(0.013)	(0.082)	(0.029)
Treatment*high social desirability	1.51	-0.0095	0.11	-0.010
	(2.13)	(0.016)	(0.096)	(0.035)
Control mean	36.77	0.17	0.36	0.28
N	1102	1102	246	1102
Controls	Lasso	Lasso	Lasso	Lasso
Sample	Men	Men	Men	Men

Notes: Robust standard errors in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Table 7: Social desirability. Women

	(1)	(2)	(3)	(4)
	Belief	Own attitude	Plan	Vignette
Information treatment	0.53	0.0074	0.042	-0.010
	(1.56)	(0.0072)	(0.069)	(0.026)
High social desirability	-2.33	0.014	-0.054	-0.074***
	(1.82)	(0.011)	(0.069)	(0.026)
Treatment*high social desirability	-0.93	-0.025**	-0.035	0.017
	(2.13)	(0.012)	(0.080)	(0.031)
Control mean	34.43	0.12	0.31	0.23
N	1242	1242	274	1242
Controls	Lasso	Lasso	Lasso	Lasso
Sample	Women	Women	Women	Women

Notes: Robust standard errors in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Overall, these results indicate that while baseline social desirability is strongly related

to the level of reported attitudes and intentions about FGM, it does not appear to differentially shape the response to the information intervention. High- and low-social-desirability respondents react in broadly similar ways to the treatment. We therefore interpret our main experimental findings as unlikely to be driven by differential social desirability bias.[4]

5 Discussion

In this section, we briefly discuss reporting changes over time and why our experiment only affects men.

5.1 Massive changes over time

Before turning to mechanisms, it is useful to step back and consider how attitudes, beliefs, and intentions evolve between the baseline and the follow-up. Figure 4 plots the mean of our four primary outcomes at wave 1 (baseline) and wave 2 (follow-up) for the experimental sample. The changes are large.

[4]In the pre-analysis plan, we also stated that we would look at (i) the potential size of belief updating, (ii) heterogeneous treatment effects using an honest causal forest approach, and (iii) an IV regression using beliefs as an instrument for the treatment. Given our results, we have decided not to move forward with these analyses.

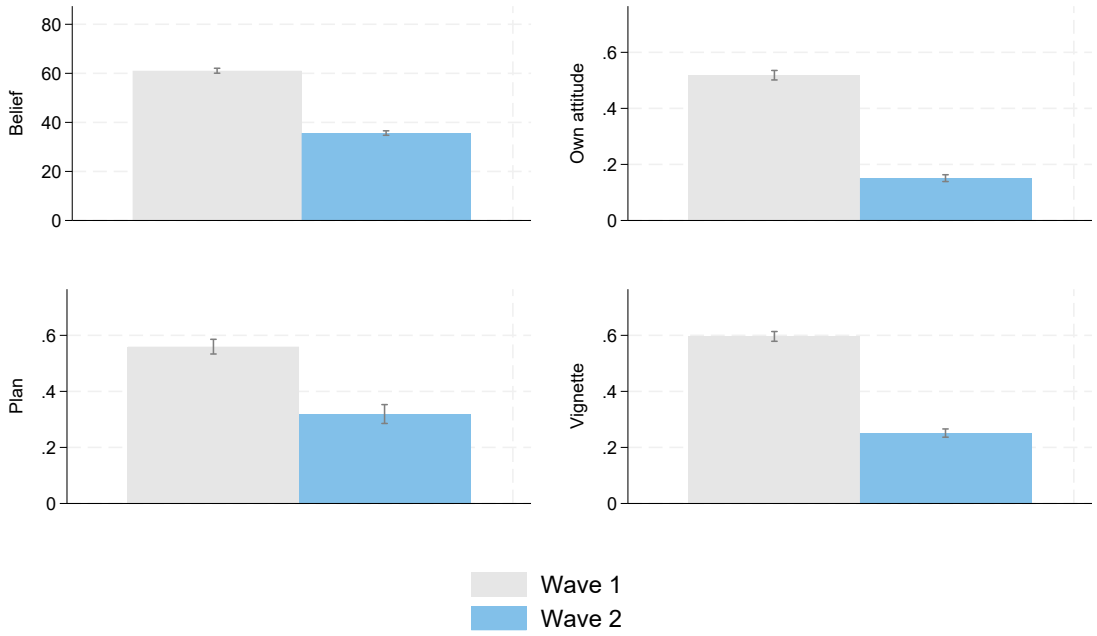


Figure 4: Changes in outcomes over time.
The sample consists of all mothers and fathers in the experimental sample.

Between the two waves, average beliefs about others' support for FGM fall from 61% to about 36%, and the share of respondents who agree that "a girl in our village should be circumcised" declines from 52% to roughly 15%. The proportion who report that they would cut a hypothetical daughter in the vignette drops from 60% to 25%, while plans to cut an existing daughter also decline substantially. These changes occur for both men and women and for both treated and control respondents.

Such large shifts are striking. One interpretation is that they reflect genuine and rapid changes in social norms and individual preferences, driven by a combination of national policy, NGO campaigns, religious messaging, and broader public debate around FGM during this period. Another, not mutually exclusive, interpretation is that they partly reflect changes in reporting: as opposition to FGM becomes more salient and more strongly endorsed by authorities, respondents may become less willing to voice support for the practice in a survey. The increase we observe in the Marlowe-Crowne social desirability index between

baseline and follow-up is consistent with this latter interpretation (see Figure A.4).

At the same time, several pieces of evidence suggest that the entire pattern cannot be explained by social desirability alone. First, when we focus on respondents who have low social desirability at follow-up, we still see substantial declines in beliefs, attitudes, and intentions between waves (see Figure A.5). Second, as shown in Section 4.3, baseline social desirability moderates the *level* of reported intentions but does not systematically change how respondents react to the information treatment. Third, when we exclude respondents who report at least one cut daughter at baseline but no cut daughters at follow-up, our main treatment effects are essentially unchanged (see Appendix Table A.1).

Importantly, despite these large changes in levels, the key baseline patterns that motivated our experiment remain clearly visible at follow-up. When we replicate the three baseline "facts" using the follow-up data, we find that (i) many parents who say that girls should not be circumcised still have a cut daughter or plan to cut a daughter, (ii) parents continue to overestimate local support for FGM, and (iii) misperceptions of others' attitudes are still strongly positively correlated with intentions to cut (Figures A.6–A.8). In other words, the entire distribution has shifted in a more progressive direction, but the underlying relationship between misperceived norms and intended cutting decisions remains in place.

Given the size of the changes between baseline and follow-up, we are cautious about interpreting them as pure shifts in underlying preferences or social norms. They are likely to reflect some combination of genuine change and changes in reporting behavior, and our data do not allow us to cleanly separate the two. This also implies a limitation for our experimental design. The information statistic X is constructed from baseline data, and our sample restrictions guarantee that, at baseline, the information we provide is always (weakly) more progressive than respondents' beliefs and than the situation in their own kebele. If attitudes and beliefs have in fact become substantially more progressive by the time of the follow-up, however, the same baseline-based statistic may no longer be strictly

progressive relative to respondents' updated beliefs or to the current environment. In the extreme, for some respondents the information could even suggest that more people support FGM than they would otherwise have thought. We see no evidence in our data that the intervention increased support for FGM or intentions to cut daughters, but this possibility highlights that our estimates capture the marginal impact of a one-off information treatment based on past data in a rapidly evolving context, rather than the effect of providing fully up-to-date information about prevailing norms.

More broadly, the combination of large changes in reported attitudes and intentions and our limited ability to disentangle genuine change from shifts in reporting points to the need for more consistent monitoring of FGM practices. In most settings, including ours, policymakers and researchers rely on infrequent household surveys with self-reported cutting status and attitudes. Our results suggest that substantial progress may be occurring, but without regular and well-designed data on both actual cutting and the social environment in which decisions are made, it is difficult to distinguish true behavioral change from changing survey responses. Investing in better and more frequent monitoring of FGM, including innovations in measurement where feasible, would allow future work to track trends more reliably and to evaluate interventions against a clearer picture of underlying dynamics.

5.2 Why do effects appear for men but not for women?

A striking feature of our results is that the information treatment reduces men's stated intentions to cut their daughters, while we find no detectable effects for women. At the same time, the treatment leaves both men's and women's reported beliefs about others' support for FGM and their own attitudes essentially unchanged. This pattern raises the question of why men, but not women, adjust their stated intended behavior in response to information about opposition to FGM in nearby communities.

Our data do not allow us to identify mechanisms conclusively, so the discussion here is

necessarily speculative. We highlight two possible, non-mutually exclusive explanations.

First, men and women may face different constraints when acting on their preferences. Qualitative work and survey evidence in other settings suggest that women, and especially mothers and grandmothers, are often responsible for organising the cutting itself, but that decisions are embedded in a wider family and community structure in which male approval or veto power matters.[5] If women anticipate stronger sanctions from other women or from extended family for deviating from prevailing practices, they may be less willing or able to change their stated plans for daughters in response to information about others' opposition. Men, in contrast, may perceive more room to adjust what they say they intend to do when confronted with evidence that others are more opposed than they thought.

Second, there may be gender differences in how respondents map information about community norms into survey responses. It is possible that men view our vignette and plan questions as closer to hypotheticals, whereas women, who are more involved in the practical arrangements around cutting, may interpret them as commitments that are harder to revise in a survey context. If so, men might be more willing to adapt or change these stated intentions when they learn that others are more opposed than they had believed, even without changing their reported beliefs or attitudes on the survey scales.

We emphasize that these interpretations remain tentative. The information intervention is deliberately light-touch, and the effects on men's intentions are modest in size. Nonetheless, the fact that we observe any change in stated intentions for men, despite no detectable shift in reported beliefs or own attitudes and no effects for women, suggests that norm-based information may interact in complex ways with gender-specific roles and constraints within the household.

A further limitation of our design is that mothers and fathers were interviewed and treated

[5]For example, Ferreira et al. (2024) report that in Somalia, a large majority of families attribute the decision to cut to the mother, but decisions often take place in consultation with other family members.

individually during the same survey visit, and our main outcomes were measured immediately after the information script. In reality, decisions about cutting are often discussed within the couple and with extended family members over time. By construction, our experiment leaves little scope for spouses to process the information together, renegotiate plans, or coordinate on a joint response. Future interventions could explicitly build in time and structure for intra-household discussion—for instance, by providing both spouses with the same information, offering discussion prompts, and measuring intentions and behavior after a delay. Such designs would be better suited to capturing whether norm information can shift shared decisions about cutting, rather than only individual survey responses. Designing interventions that more directly target these constraints, and that can follow actual cutting decisions over a longer period, remains an important task for future research.

6 Conclusion

Female genital mutilation remains highly prevalent in many countries and continues to affect millions of girls. Understanding why parents persist in cutting their daughters, and how policy can support abandonment, requires evidence on the social and intra-household mechanisms that sustain the practice. In this paper, we combine rich data from Ethiopia with a pre-registered information experiment to study the role of misperceived social norms and how mothers and fathers respond to norm-based information.

The baseline data reveal three robust patterns. First, even among parents who state that girls in their village should not be circumcised, many have at least one cut daughter and a substantial share still plan to cut an uncut daughter or say they would cut a hypothetical daughter. Second, parents substantially overestimate local support for FGM: they believe that many more women and men in their communities support cutting than we observe in the same kebeles. Third, these misperceptions are strongly positively correlated with intended

cutting. These facts motivate our experimental intervention, in which we provide parents with accurate information about opposition to FGM in a nearby community.

The information treatment leaves reported beliefs about others' attitudes and respondents' own attitudes essentially unchanged. The null effects on beliefs and attitudes are estimated precisely enough that we can rule out moderately large average changes. At the same time, we find some evidence that the intervention affects stated intentions for men: fathers who receive the information are less likely to report that they plan to circumcise a daughter and less likely to say they would circumcise a hypothetical daughter. The corresponding reductions are on the order of 20–25 percent and 10–15 percent of the control means, respectively, although statistical significance is sensitive to corrections for multiple hypothesis testing. For mothers, we do not detect effects on beliefs, attitudes, or intended cutting. Disaggregating by information type, we do not find systematic differences between learning about women's versus men's opposition to FGM in a nearby community, and baseline social desirability does not appear to systematically moderate the impact of the treatment.

For policy, our findings suggest that norm-based information can play a useful, if limited, role alongside more intensive community engagement and empowerment programs. Our experiment shows that providing men with credible evidence that others oppose FGM can reduce stated intentions to cut, even in a context of rapidly changing norms and without large shifts in reported attitudes. In practice, similar information could be incorporated as a low-cost component of broader initiatives—for example, by feeding back up-to-date, locally collected statistics on opposition to FGM in community meetings or in couple-focused counselling, with particular attention to engaging fathers. At the same time, our experience underscores the importance of ensuring that such information is based on current data and that it is clearly more progressive than what participants already believe, to avoid the risk of inadvertently reinforcing support for the practice.

Future research that combines norm information with interventions that more directly

target decision-making power within households and communities, and that tracks actual cutting outcomes over a longer horizon, will be important to assess whether and how information about social norms can contribute to the sustained abandonment of FGM.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT (OpenAI) to assist with improving the clarity of the writing, suggesting alternative phrasings, and helping to structure parts of the introduction, discussion, and conclusion. No AI tools were used to generate, analyze, or interpret the data or results. After using this tool, the authors reviewed and edited all content as needed and take full responsibility for the content of the published article.

References

Alemu, Addisu Alehegn. 2021. "Trends and Determinants of Female Genital Mutilation in Ethiopia: Multilevel Analysis of 2000, 2005 and 2016 Ethiopian Demographic and Health Surveys." International Journal of Women's Health 13:19.

Belloni, Alexandre, Victor Chernozhukov and Christian Hansen. 2014. "Inference on treatment effects after selection among high-dimensional controls." Review of Economic Studies 81(2):608–650.

Blair, Graeme, Alexander Coppock and Margaret Moor. 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.

Bursztyn, Leonardo, Alessandra L González and David Yanagizawa-Drott. 2020. "Misperceived social norms: Women working outside the home in saudi arabia." American economic review 110(10):2997–3029.

Camilotti, Giulia. 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.

Central Statistical Agency (CSA) [Ethiopia] and ICF. 2016. Ethiopia Demographic and Health Survey 2016. Technical report Central Statistical Agency (CSA) [Ethiopia] and ICF Addis Ababa, Ethiopia and Rockville, Maryland, USA: .

Congdon Fors, Heather, Ann Sofie Isaksson and Annika Lindskog. 2021. "Harmful norms: Can social convention theory explain the persistence of female genital cutting in Africa?" Mimeo Gothenburg Univeristy .

Crowne, Douglas P and David Marlowe. 1960. "A new scale of social desirability independent of psychopathology." Journal of consulting psychology 24(4):349.

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

Dhar, Diva, Tarun Jain and Seema Jayachandran. 2018. "Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India.".

Efferson, Charles, Sonja Vogt, Amy Elhadi, Hilal El Fadil Ahmed and Ernst Fehr. 2015. "Female genital cutting is not a social coordination norm." Science 349(6255):1446–1447.

Elmusharaf, Susan, Nagla Elhadi and Lars Almroth. 2006. "Reliability of self reported form of female genital mutilation and WHO classification: cross sectional study." Bmj 333(7559):124.

Ferreira, Pedro, Selim Gulesci, Eliana La Ferrara, David Smerdon and Munshi Sulaiman. 2024. "Changing Harmful Norms through Information and Coordination: Experimental Evidence from Somalia." Preliminary draft, 25 October 2024.

Fink, Günther, Margaret McConnell and Sebastian Vollmer. 2014. "Testing for heterogeneous treatment effects in experimental data: false discovery risks and correction procedures." Journal of Development Effectiveness 6(1):44–57.

Fors, Heather Congdon, Ann-Sofie Isaksson and Annika Lindskog. 2024. "Changing local customs: The long run impacts of Christian missions on female genital cutting in Africa." Journal of Development Economics 166:103180.

Gibson, Mhairi A., Eshetu Gurmu, Regina Chua, Hannelore Van Bavel and Sarah Myers. 2023. "Abandoning female genital mutilation/cutting (FGMC) is an emerging but costly

parental investment strategy in rural Ethiopia." Social Science Medicine 335:116170. **URL:** *https://doi.org/10.1016/j.socscimed.2023.116170*

Glynn, Adam 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, Selim, Sam Jindani, Eliana La Ferrara, David Smerdon, Munshi Sulaiman and H Young. 2021. "A Stepping Stone Approach to Understanding Harmful Norms.".

Haaland, Ingar, Christopher Roth and Johannes Wohlfart. 2023. "Designing information provision experiments." Journal of economic literature 61(1):3–40.

Kotsadam, Andreas and Mette Løvgren. 2025. "Is it Time to Put a Moratorium on List Experiments for Domestic Violence Elicitation?" Sociological Methods & Research p. 00491241251372542.

Mackie, Gerry. 1996. "Ending footbinding and infibulation: A convention account." American sociological review pp. 999–1017.

Novak, Lindsey. 2020. "Persistent norms and tipping points: The case of female genital cutting." Journal of Economic Behavior & Organization 177:433–474.

Reynolds, William M. 1982. "Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale." Journal of clinical psychology 38(1):119–125.

Shell-Duncan, Bettina, Katherine Wander, Ylva Hernlund and Amadou Moreau. 2011. "Dynamics of change in the practice of female genital cutting in Senegambia: testing predictions of social convention theory." Social science & medicine 73(8):1275–1283.

UNFPA and UNICEF. 2024. "UNFPA–UNICEF Joint Programme on the Elimination of Female Genital Mutilation." https://www.unicef.org/protection/unfpa-unicef-joint-programme-eliminating-fgm. Accessed 1 December 2025.

UNICEF. 2024. "Female Genital Mutilation: A global concern. 2024 Update.".

URL: *https://data.unicef.org/wp-content/uploads/2024/03/FGM-Data-Brochure-v13.4.pdf*

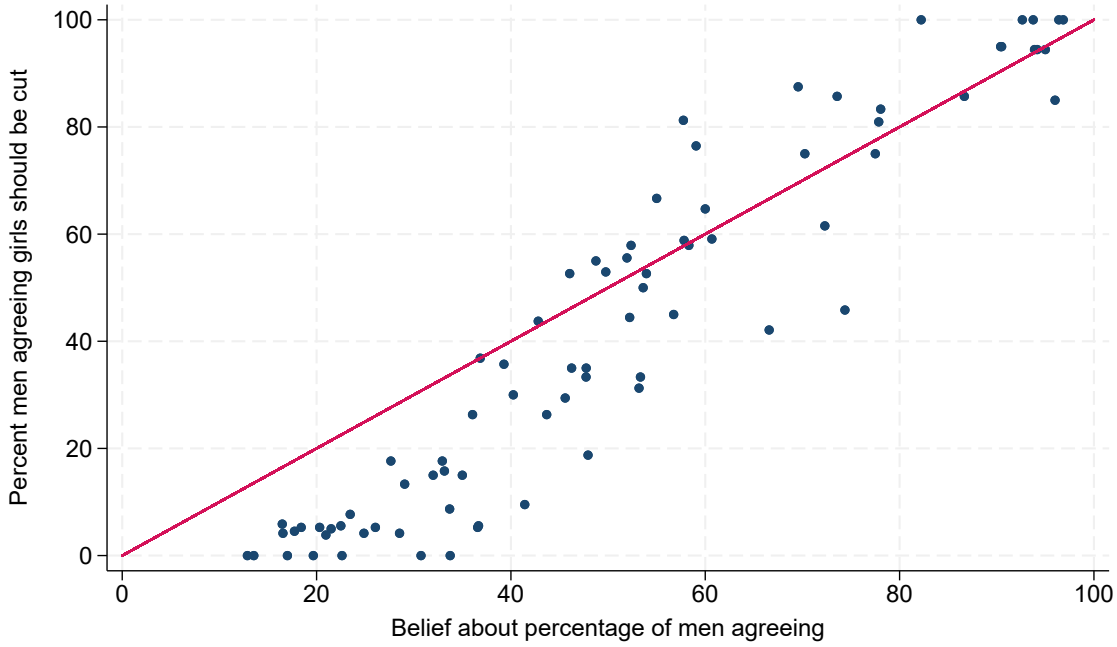
UNICEF Innocenti Research Centre. 2005. "Changing a Harmful Social Convention: Female Genital Mutilation/Cutting.".

United Nations Population Fund (UNFPA). 2025. "Female genital mutilation." https://www.unfpa.org/female-genital-mutilation. Accessed 18 November 2025.

Appendix:

A.1 Additional tables and figures

Figure A.1: Observation 2: Men tend to overestimate the support of FGM in their community



Notes: Kebele level averages. Full sample of men at baseline. 45-degree line in red.

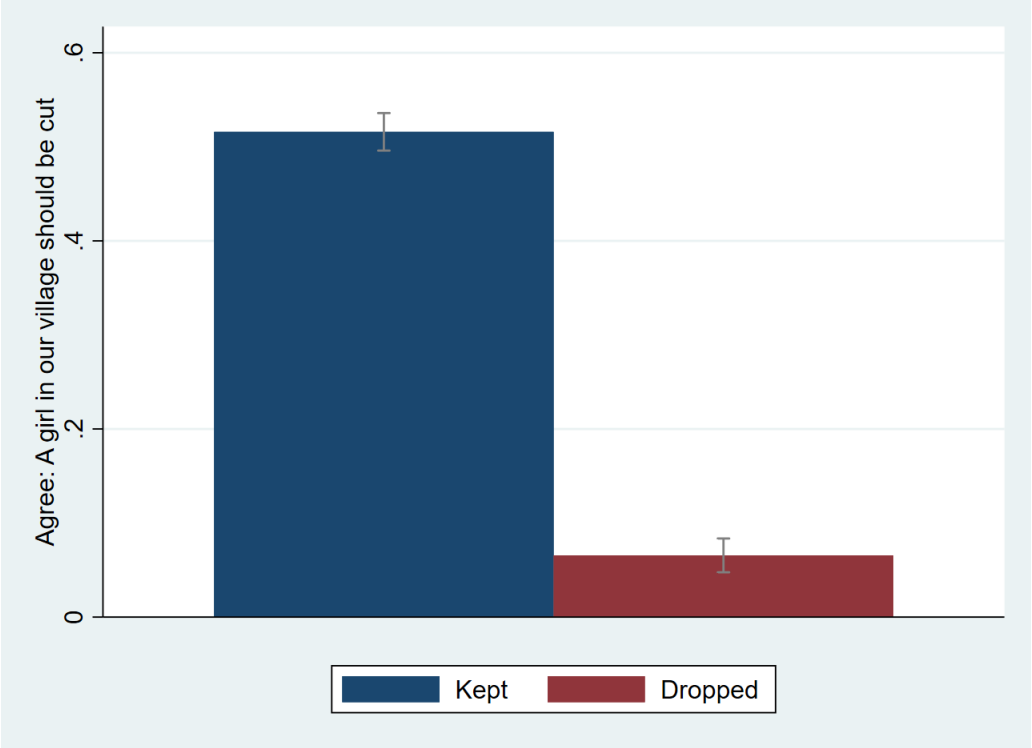


Figure A.2: Individual attitudes in kept and dropped sample

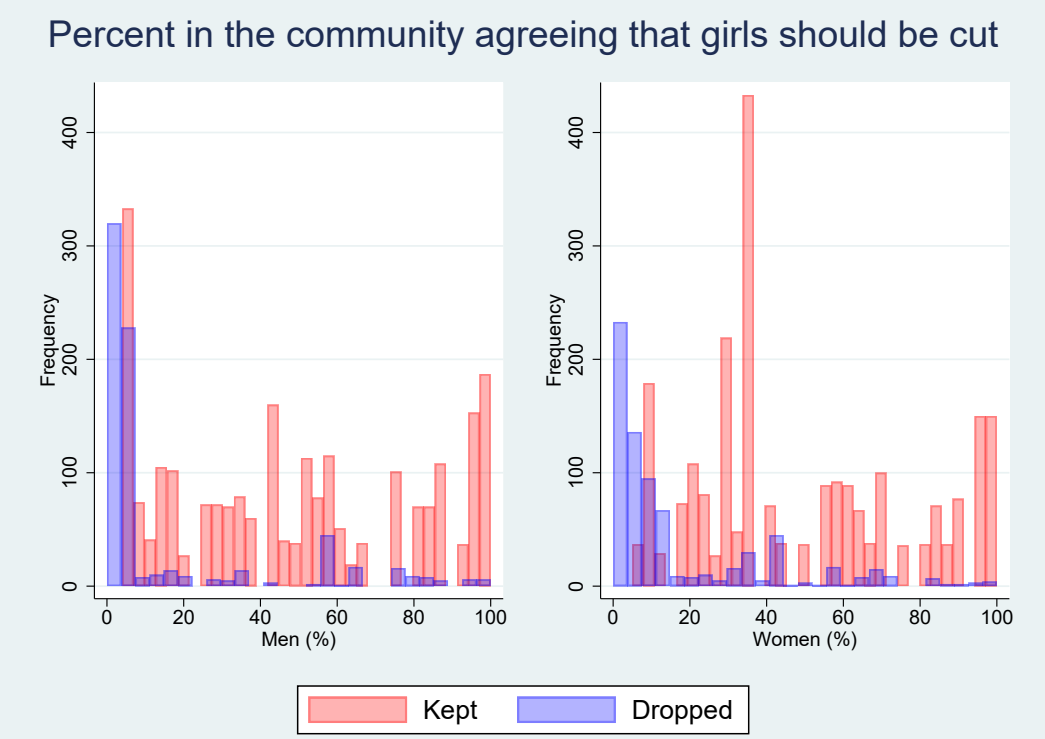


Figure A.3: Community attitudes in kept and dropped sample

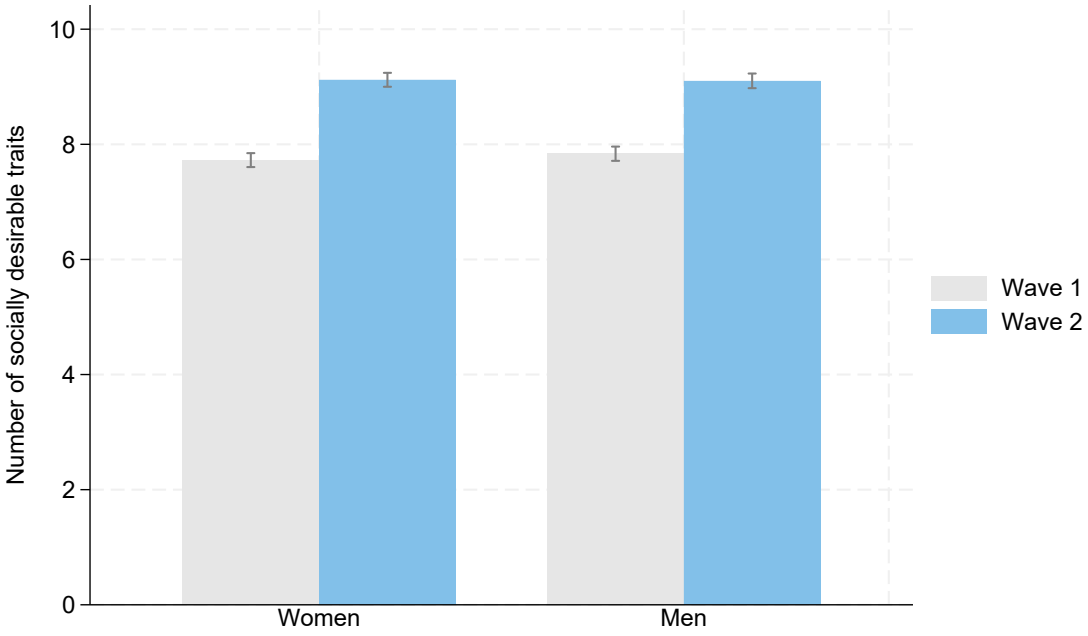


Figure A.4: Changes in social desirability over time.
The sample consists of all mothers and fathers in the experimental sample.

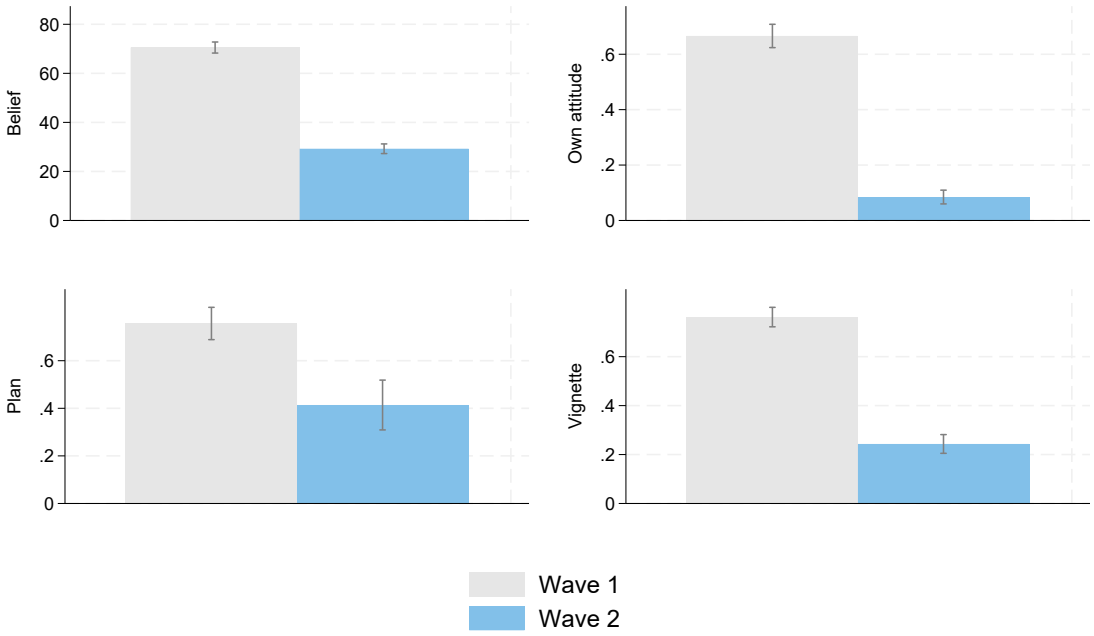


Figure A.5: Changes in outcomes over time. Low social desirability

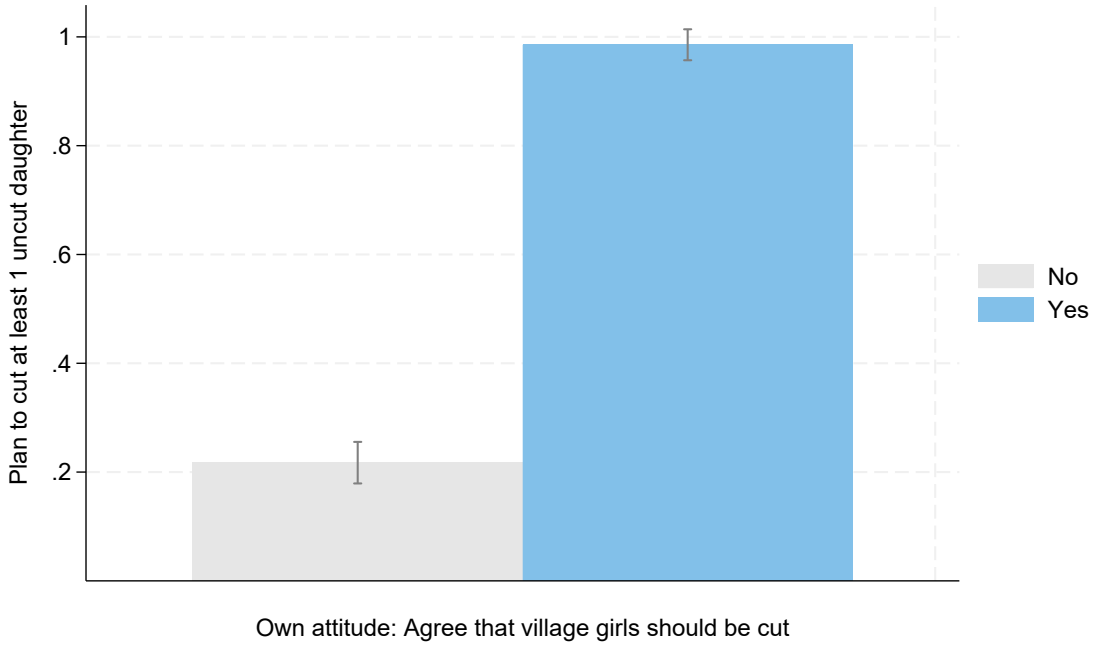
The sample consists of mothers and fathers in the experimental sample with low social desirability at endline.

Table A.1: Drop the 9 percent liars, men

	(1)	(2)	(3)	(4)
	Belief	Own attitude	Plan	Vignette
Information treatment	-1.44	-0.0046	-0.055	-0.037**
	(1.08)	(0.0073)	(0.042)	(0.017)
Control mean	36.86	0.18	0.38	0.29
N	1035	1035	203	1035
Controls	Lasso	Lasso	Lasso	Lasso
Sample	Men	Men	Men	Men

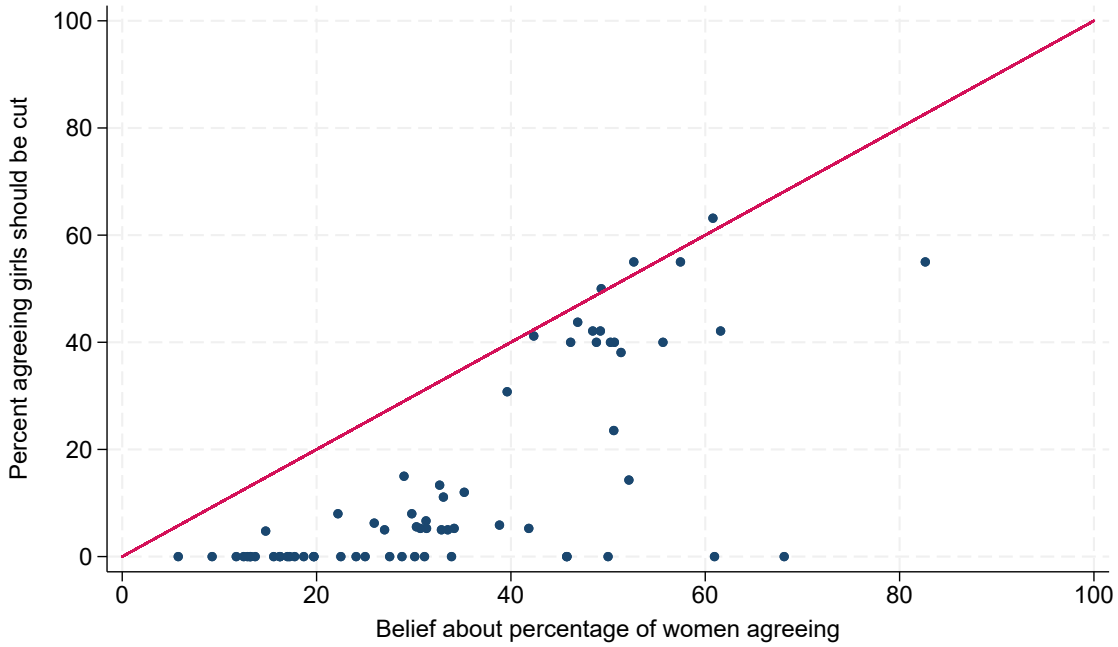
Notes: Robust standard errors in parentheses. P-values are $\leq 0.01^{***}$, $\leq 0.05^{**}$, and $\leq 0.1^{*}$.

Figure A.6: Observation 1: Many parents cut their daughters even if they oppose the practice



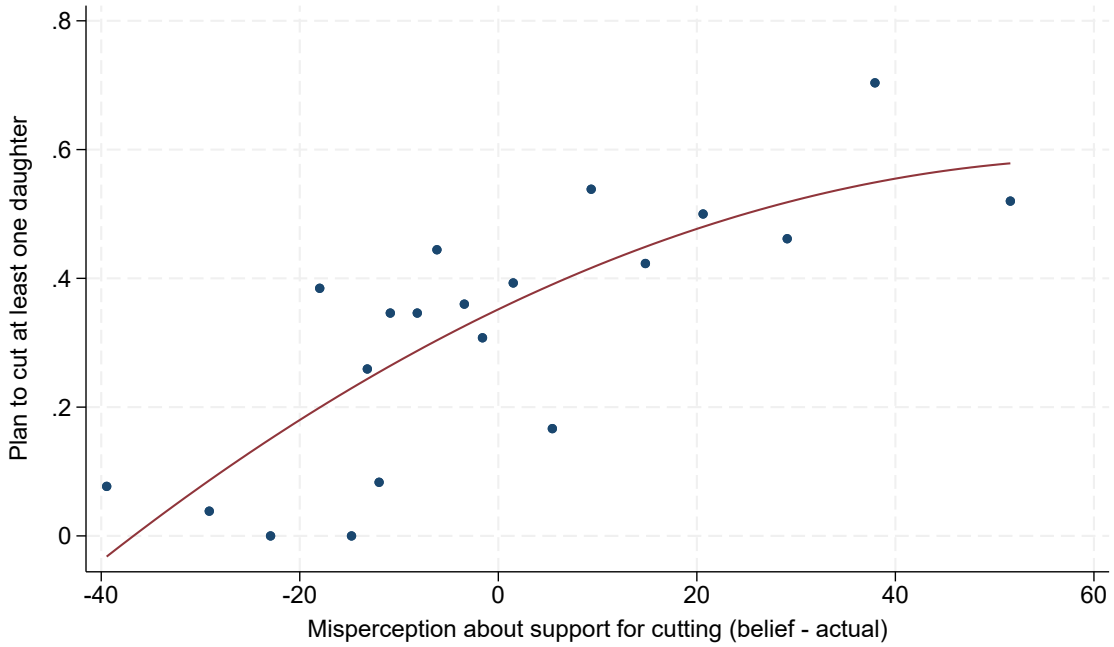
Notes: Experimental sample in follow up data for cutting, only those with at least one uncut daughter for plans. 95 percent confidence intervals.

Figure A.7: Observation 2: Parents tend to overestimate the support of FGM in their community



Notes: Kebele level averages. Experimental sample in follow up data. 45-degree line in red.

Figure A.8: Observation 3: There is a strong correlation between misperceptions and intention to cut a daughter in the future



Notes: Experimental sample in follow up data. The misperceptions are calculated as beliefs about attitudes of women minus actual attitudes, for women the actual attitudes removes the respondent herself.

A.2 Reporting: Social desirability and list experiment

An explanation for the observed discrepancy between attitudes and behavior could be misreporting. Self-reported attitudes and behaviors may suffer from a social desirability bias if people are ashamed of their behavior or if they feel that the enumerators may think less of them if they have certain attitudes. Such worries are largest with sensitive topics. In this section, we will investigate to what extent there seems to be a social desirability bias in our data. We do this by first using a social desirability index. We then assess bias in reporting of circumcised daughters by comparing reporting of the same phenomena for fathers and mothers in the same household. We then move on to conducting a list experiment to assess

bias in reported attitudes. In this section, we use the survey responses collected from the parents and the level of analysis is the parents.

A.2.1 Social desirability scale

To investigate social desirability in our survey, we use the Marlowe-Crowne social desirability scale (Crowne and Marlowe, 1960). We use a shorter survey module developed by Reynolds (1982) which consists of asking the respondents whether they possess 13 very good traits, such as always being a good listener or always admitting making a mistake. The traits are sometimes reverse-coded to avoid capturing people always answering the same to be coded as having a high social desirability. It is very unlikely that a person actually has all 13 saintlike traits. Figure A.9 shows the distribution of the traits in our sample. We see that less than 1 percent of the sample possess all 13 traits, but 22 percent possess at least 10 of them. The median number of traits is between 7 and 8, and following Dhar, Jain and Jayachandran (2018) we code individuals with 8 or more traits as having a high social desirability, which is indicated by the dashed line in the figure.

In Table A.2 we see that individuals with a high social desirability score report having fewer cut daughters (column 1) and that they have a 17 percentage point lower probability of reporting to have at least one cut daughter (column 3). There may of course be other differences between individuals with different degrees of social desirability, and we see that the correlations are lower when we add control variables in columns 2 and 4. We therefore move on to investigate reporting also with other methods.

A.2.2 Within household differences in reporting

Mothers and fathers are interviewed separately in our survey. Since the survey was unannounced beforehand, there is no reason to expect that spouses colluded to decide what to

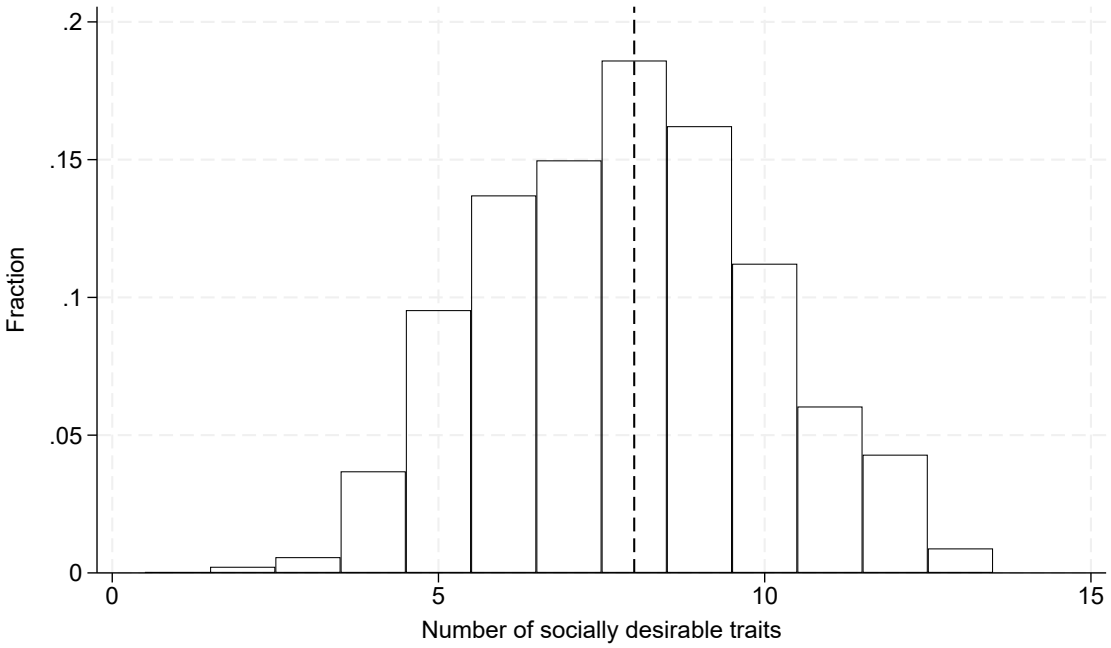


Figure A.9: Distribution of traits in the social desirability scale.
The sample consists of all mothers and fathers.

Table A.2: Correlation between reporting circumcision and high social desirability.

	(1) Number of cut daughters	(2) Number of cut daughters	(3) Any cut daughter	(4) Any cut daughter
High social desirability	-0.29***	-0.082***	-0.17***	-0.038***
	(0.038)	(0.028)	(0.017)	(0.014)
Mean outcome in low group	1.18	1.18	0.70	0.70
No. of observations	3145	3120	3145	3120
Indiv. level controls	No	Lasso	No	Lasso

Notes: The sample consists of all mothers and fathers.

answer. We can thereby test whether the fathers and the mothers report the same number of daughters that are circumcised. If we find a difference it could be because one of the parents just misremember or actually do not know. But if they answer the same number of daughters and this number is above zero, this is a good indication of there being no bias in the reporting. In Figure A.10 we see that over 97 percent of the individuals answer exactly the same as their partner.

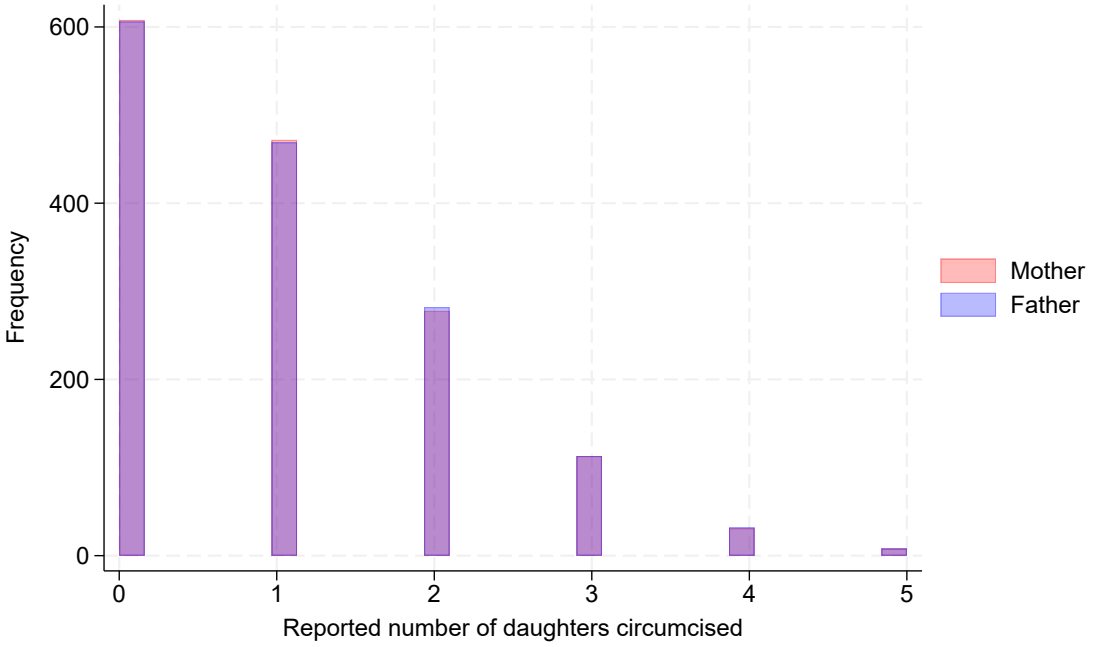


Figure A.10: Distribution of reported number of circumcised daughters.
The sample consists of all mothers and fathers in households where both were interviewed.

We assume that couples answering the exact same positive number report the correct one. With individuals answering differently, we may be worried that the one answering the lower number is misrepresenting reality. It is also possible that couples where both answer zero daughters partly consist of individuals where both misrepresent reality.

A.2.3 Assessing reporting issues: results from list experiments

As a final test to assess whether there is underreporting and social desirability bias in our data, we use a so-called "list experiment" (also known as the "item count technique"). The list experiment is conducted by randomly splitting the sample into two groups. The individuals are then asked to count the number of true statements on a list that either includes a sensitive statement or not. By comparing the number of statements reported as true across the two groups we get a measure without any specific individual having revealed their own status. By also asking a question about the sensitive statement directly to the respondents we can assess the degree of underreporting by comparing the results when using the two different ways of asking.

The list experiment was conducted at the end of the survey. First, the enumerator clearly explained the procedure with the help of an example and explained to the participants that their individual answers could not be detected with this approach (see Figure A.11). In Figure A.12, we show the control and treatment questions when the variable of interest is "A girl should be circumcised". The control questions include four statements that we are not interested in and that are used only to get an average to compare the other group with. The treatment list includes the same questions and adds the question of interest. The control questions are created to avoid ceiling and floor effects and to include items that are negatively correlated so as to increase power (Glynn, 2013). To take a concrete example, let us say that the control group answers that two of the four statements are true on average and the treatment group answers that 2.3 of the statements are true on average. Since the only difference between the two groups are the extra question on circumcision we would infer that 30 percent of the individuals in the list treatment group think that "A girl should be circumcised".[6]

[6]We also included a second list experiment where we swapped the treatment and control groups but due to translation errors we can not use this experiment.

"We have some questions about your life and your attitudes, and I will now explain how you should respond to these.

Now I'm going to read some statements about many different things. Some of these statements will be true and some will not. After I read all statements, please tell me HOW MANY of them are true for you.

And this is important: I do not want to know which ones, just how many.

Let's try with an example first. Suppose I read you 4 statements. After I read each statement, I want you to count with your fingers if it is true, and keep track **without showing to me**. I will then ask you how many are true. Let me demonstrate first.

NOTE TO ENUMERATOR: READ THE STATEMENTS BELOW AND SHOW THEM HOW YOU ARE COUNTING:

1. I went to the movies with my best friend on Sunday
2. I had dinner last night
3. I can speak English
4. Addis Ababa is the capital of Ethiopia

I will now read you the statements. As I'm reading, count with your fingers without showing me (**ENUMERATOR LOOK THE OTHER WAY**). At the end, you'll tell me how many are true for you.

1. You went to the movies with my best friend on Sunday
2. You had dinner last night
3. You can speak English
4. Addis Ababa is the capital of Ethiopia

How many of these statements are true?

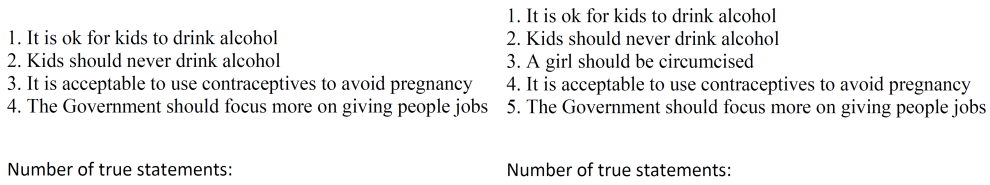
Note that I do not know which of the statements that are true if you just give me the number. I only know how many. In this module, you will have several examples of this, remember to only tell us how many things are true so that we can not know which ones are true.

Figure A.11: Instructions for the list experiment

Table A.3: List experiment.

	(1) Number of true statements	(2) Number of true statements
List treatment	0.38***	0.37***
	(0.027)	(0.026)
Mean nr answers in C group	2.27	2.27
Mean direct question in C group	0.41	0.41
No. of observations	3145	3120
Indiv. level controls	No	Lasso

Notes: The List treatment group get the statement "A girl should be circumcised". Robust SE in parentheses.



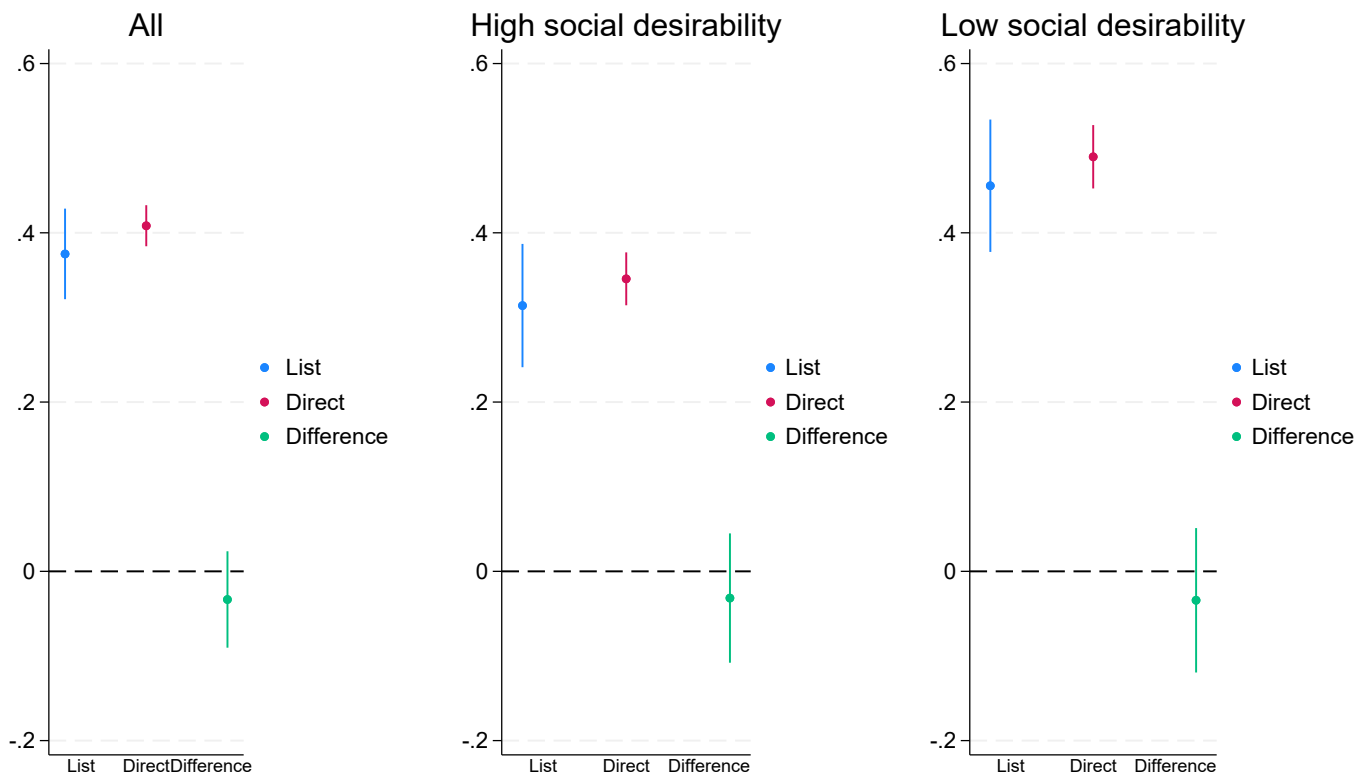
(a) Questions to the list experiment control group.

(b) Questions to the list experiment treatment group.

Figure A.12: List experiment for the question "A girl should be circumcised"

In Table A.3, we show the results of the list experiments. We see that individuals in the control group answers that 2.27 of the statements are true on average. This average is 0.38 higher in the treatment group, with the interpretation that 38 percent of the treatment group think that a girl should be circumcised. This can be compared to when asking the question directly where we see that 41 percent of the responders in the control group answer that they think that a girl should be circumcised. Based on this, there is no sign of biased reporting overall, and the list experiment even yields a lower number than asking directly (see Kotsadam and Løvgren (2025) for a general discussion of this "fleeing behavior" in list experiments). The difference is not statistically significant which can be seen in the green dot to the left in Figure A.13.

A major disadvantage of list experiments in many settings is that they suffer from much higher variability than direct questions. Blair, Coppock and Moor (2020) show that list experiments are typically 14 times noisier than direct questions and that most studies, which



Notes: List refers to the estimated prevalence of people thinking "A girl should be circumcised" in the list experiment. Direct refers to the prevalence when using a direct survey question. Difference refers to the difference between asking in the list experiment minus asking directly. 95 percent confidence intervals are shown.

Figure A.13: List experiment: "A girl should be circumcised" by sub-groups

usually have less than 2000 observations, are severely underpowered to use them. For example, De Cao and Lutz (2018) use a list experiment in Ethiopia and find 10 percentage points more support for FGM than in direct questioning. However, their sample consist of less than 800 individuals and they cannot reject that the bias is even in the other direction or that it is larger than 20 percentage points. In contrast, we can reject even small differences in the different types of questioning.

The degree of underreporting can also be compared across subgroups and in Figure A.13 we split the samples into those with a high and low social desirability scale. We again note that the individuals with a high score on the social desirability scale report lower numbers of cut daughters but there is no difference between the list and direct reporting for any of the groups.

A.3 Coding of variables

A.3.1 Primary and Secondary Dependent Variables

Our short-run outcomes, Y, are:

- **Decision about FGM for a hypothetical girl (Vignette):** If you have a daughter in the future, do you intend to circumcise her? The variable is an indicator variable taking the value 1 if the respondent answers "yes," and 0 if the respondent answers "no."

- **Own attitudes about FGM:** Do you agree or disagree with the following statement: *"A girl in our village should be circumcised."* Indicator variable taking the value 1 if the respondent agrees and 0 if the respondent disagrees. This variable is missing for everyone that do not have daughters or who only have cut daughters.

- **Plans for uncut daughters, where applicable:** Do you plan to circumcise [name

of daughter]? The variable is an indicator variable taking the value 1 if the respondent answers "yes" for at least one daughter, and 0 if the respondent answers "no" for all daughters.

- **Beliefs about attitudes towards FGM in the community:** The average of the answers to the two questions: *Imagine that there are 100 women in your village. Of these, how many agree with the statement: "A girl in our village should be circumcised?"* Discrete variable between 0 and 100. And: *Imagine that there are 100 men in your village. Of these, how many agree with the statement: "A girl in our village should be circumcised?"* Discrete variable between 0 and 100.

These four variables are our main variables of interest.

A.3.1.1 Secondary Outcome: Social Desirability Index

We create a Social Desirability Index based on responses to a Marlowe-Crowne module, which measures a person's general tendency to give socially desirable answers. The index is based on the following questions where individuals can answer "agree" or "disagree":

1. It is sometimes hard for me to go on with my work if I am not encouraged.
2. I sometimes feel resentful when I don't get my way.
3. On a few occasions, I have given up doing something because I thought too little of my ability.
4. There have been times when I felt like rebelling against people in positions of authority even though I knew they were right.
5. No matter who I'm talking to, I'm always a good listener.
6. There have been occasions when I took advantage of someone.

7. I'm always willing to admit it when I make a mistake.

8. I sometimes try to get even rather than forgive and forget.

9. I am always courteous, even to people who are disagreeable.

10. I have never been irked when people expressed ideas very different from my own.

11. There have been times when I was quite jealous of the good fortune of others.

12. I am sometimes irritated by people who ask favors of me.

13. I have never deliberately said something that hurt someone's feelings.

We use the baseline answers to these questions and create a variable *High Social Desirability*, which equals 1 if the respondents answer more than 7 of the questions in a socially desirable way. We test for treatment effect heterogeneity by interacting this variable with treatment. The worrisome pattern would be if the treatment effects were driven by individuals with a high propensity to disingenuously give socially desirable answers and vanished for those with a low such tendency.

A.3.2 Main Independent Variables

Our main treatment variable is: *Information about high share of FGM opposition*. This variable is coded as 1 if information about either women's or men's high share of opposition is given, and 0 otherwise.

The other main independent variables are all from the baseline survey, and they are:

- Continuous numerical values of:
  - Age
  - Education (*"What is the highest level of education you have completed?"*)

  - Number of children
  - Number of girls
  - *"How common is it currently for girls to be circumcised in your community?"*
  - *"From what you know, how does your religion view female circumcision?"*

- Beliefs about:
  - Percentage of men in the village who agree with FGM for girls
  - Percentage of women in the village who agree with FGM for girls

- Dummy variables for:
  - Employment status (1 if having engaged in income-generating activities)
  - Literacy (1 if yes on *"Do you know how to read and write in any language?"*)
  - Gender of respondent
  - Answering "Yes" to *"If you have a daughter in the future, do you intend to circumcise her?"*
  - Thinking that *"A girl in our village should be circumcised"*

We also add the following pre-treatment questions from the midline survey:

- Dummy variables for:
  - *Dummy for stopped: "Do you think that female circumcision should be continued, or should it be stopped?"* (1 = continued, 2 = stopped, 3 = do not know)
  - *"Do you think that FGM is harmful to the girl?"* (1 = Yes, 2 = No)
  - *"Do you think FGM is violating the rights of the girl?"* (1 = Yes, 2 = No)

- Continuous numerical values of:

- *To what extent do you agree that...* Response options:
  - Cutting shows respect to our elders. (1. strongly agree, 2. agree, 3. neither agree nor disagree, 4. disagree, 5. strongly disagree)
  - Cutting helps a girl stay a virgin until she marries.
  - Cutting teaches girls obedience and respect.
  - Cutting is not the right thing to do to girls in our community.
  - Cutting marks the transition from a girl child to a woman/adult.
  - Cutting is not part of our traditions and culture.
  - Cutting ensures that a girl retains her femininity.
  - Uncut girls are not pure.

Using data on over 3,000 mothers and fathers in five Ethiopian regions, we study how misperceived social norms about female genital mutilation (FGM) relate to cutting decisions and whether norm information changes them. At baseline, many parents who say FGM should stop have a cut daughter and still plan to cut, parents overestimate local support for FGM, and these misperceptions are correlated with intended cutting. Motivated by this, we field an information experiment that randomly informs parents about opposition to FGM in a nearby community. The intervention has no detectable effects on reported beliefs about others' attitudes or on respondents' own attitudes. For fathers, however, it modestly reduces stated intentions to circumcise: plans to cut a daughter fall by 8 percentage points and willingness to cut a hypothetical daughter by 4 percentage points, though these effects are only partly robust to multiple-testing adjustments. We find no effects for mothers.

Contact us

Phone: +47 93 80 00 (from Norway)
Phone: +47 55 70 55 65 (from abroad)
cmi@cmi.no
www.cmi.no
P.O. Box 6033,
N-5892 Bergen, Norway
Jekteviksbakken 31, Bergen

CMI CHR. MICHELSEN INSTITUTE

ISSN 0804-3639 (print) ISSN 1890-5048 (PDF) ISBN 978-82-8062-878-7 (print) ISBN 978-82-8062-879-4 (PDF)