# Drama and Doctrine: Lessons from an Edutainment Experiment in Egypt

Salma Mousa Ahmed Ezzeldin Mohamed Donald P. Green

March 24 2026

**Abstract**

Can narrative entertainment dislodge harmful practices — and if so, how can such changes be measured unobtrusively? We report results from a randomized edutainment intervention in Egypt, in which participants were assigned to view a state-backed drama targeting female genital mutilation (FGM) and early marriage, family planning, or a placebo storyline. Outcomes were measured through surveys and interviewer-blinded conversations six months later, analyzed at scale using large language models. The FGM drama significantly increased opposition to the practice, driven by corrections to health and religious misperceptions — corrections the state cannot reliably engineer due opposition from local clerics and doctors. Family planning effects were weak and inconsistent. These findings replicate under more challenging field conditions, among a population with deeper baseline resistance to targeted practices, boosting external validity. Interview-based measures validate and expand the constructs captured by surveys alone. Interviews reveal the spontaneous salience of FGM as a topic without showing elevated signs of respondent re-traumatization, establishing an ethical pathway for integrating open-ended measurement into experimental evaluations of harmful practices.

# 1 Introduction

On a balmy Ramadan evening in Cairo, millions of Egyptian households gather around their television sets. One show in particular, *Faten Amal Harby*, has become impossible to ignore. The drama centers on a divorced mother locked in a custody battle under personal status laws that favor men, and quickly spilled into coffee-shop debates, social media threads, and opinion editorials. Al-Azhar, the epicenter of Sunni theological thought, issued formal condemnations, citizens filed blasphemy complaints against the show's writer, and supporters praised the narrative for foregrounding women's legal precarity in Egypt (Daily News Egypt, 2022; The New Arab, 2022; Arab Gulf States Institute in Washington, 2022; Unplugged, 2023). In this, *Faten Amal Harby* was not alone. *Taht El Wesaya* (Under Guardianship), a critique of legal guardianship, pushed Members of Parliament to call for legislative review, while *Berry Leaves*, which tackled FGM, child marriage, and family planning, sparked nationwide debate (Michaelson, 2023; Unplugged, 2023; South, 2024). What millions of viewers and legislators have long intuited, that mass media can plant the seeds of cultural change, researchers increasingly confirmed. A meta-analysis of narrative entertainment experiments found that storytelling formats modestly shift audience beliefs and self-reported behaviors, often more so than non-narrative communications (Rahmani et al., 2025). The Egyptian state had a more urgent reason to authorize the drama, however: coercion had hit the point of diminishing returns.

The Egyptian government banned female genital mutilation (FGM) in 2008, the same year it set the minimum age of marriage at eighteen. In tandem with the legal changes, the state launched successive national campaigns, negotiated with Al-Azhar, and secured official religious condemnations of both practices. Yet FGM prevalence stands at 87.5% (UNICEF, 2016) — one of the highest rates in the world — and an estimated 7.1 million Egyptian women were married before the age of eighteen, including approximately 1.3 million before age fifteen (UNFPA Arab States Regional Office, 2020). In rural Upper Egypt, where FGM, child marriage, and restricted family planning are most entrenched, the reach of

Cairo's authority has always been uneven (Moustafa, 2000; Mohamed, 2024). Religious and medical institutions, even state-affiliated ones, do not speak with one voice at the village level. Instead, rank-and-file clerics and doctors openly contest the official line, continuing to legitimize and facilitate these same practices (Mohamed, 2024; Al Arabiya English, 2017; El-Dabbee et al., 2022). Indeed, medical professionals are now the most common agents conducting FGM procedures (Elnakib et al., 2025), and medical students remain poorly informed about the health complications, ethical concerns, and legal prohibitions surrounding FGM (Mostafa et al., 2006). Entrenched social norms, coupled with local authorities that the state cannot fully co-opt, work to subvert the law. As one respondent in our study puts it: *"The law will say no, but what's in [men's] heads will keep going."*

Can narrative drama succeed where the law cannot? Faced with the limits of state intervention in Egypt's rural peripheries, the United Nations gender agency, with the support of the Egyptian government, launched the 360-Degree Edutainment Campaign in 2022, featuring a mass-media drama designed to challenge FGM, child marriage, and family planning.[1] Unlike legislation or public health campaigns, narrative drama does not require the cooperation of local intermediaries to reach its audience. Such narratives are thought to shift beliefs by transporting audiences into imagined social situations, exposing them to new information and reframed norms without triggering the resistance that direct state messaging often provokes (Green and Brock, 2000; Slater and Rouner, 2002). The vast audiences that narrative entertainment attracts mean that even small but sustained persuasive effects can meaningfully affect social outcomes (La Ferrara et al., 2012a). While promising, the edutainment literature faces substantive and methodological limitations.

Substantively, the evidence base on edutainment has been dominated by public health applications like HIV prevention (Lapinski and Nwulu, 2008; Lauby et al., 2010). Far fewer

[1]In addition to *Berry Leaves*, the campaign included an interactive website, dedicated social media platforms, and community-based activities. In 2023, the campaign reached nearly 30 million people on social media and 108,000 users through the website.

evaluations address socially contested topics such as gender equality, harmful traditional practices, or family relations. This is particularly true in the Arab world, a striking gap given the region's prolific media sector and persistent gender-based harms. The evidence we do have from Egypt suggests that direct information campaigns produce little attitude change (Christia et al., 2023; Barsoum et al., 2022), consistent with the idea that the barrier to change is doctrinal resistance rather than missing information, and opening the door to exploring whether narrative entertainment can deliver messages in more emotionally resonant ways. Methodologically, outcome measurement relies almost exclusively on closed-ended surveys (Paluck et al., 2021). Qualitative depth interviews are rare, and rarer still are interviews conducted by enumerators blinded to treatment assignment. Yet blinded depth interviews offer distinctive advantages. They can capture respondents' emotional reactions, trace how intervention themes shape respondents' reasoning, and whether themes emphasized in the drama continue to resonate weeks or months later. Moreover, this method reveals whether dramas depicting sensitive topics like FGM cause distress among victimized viewers, shedding light on important ethical concerns.

We address these gaps by evaluating *Berry Leaves*, a soap opera developed in partnership with the United Nations Population Fund addressing FGM, early marriage, and family planning. Condensing the show into 1.5-hour dramas, we randomly assigned Egyptian participants in Luxor governorate to one of three conditions: (1) a drama on FGM and early marriage, (2) a drama on family planning, or (3) a placebo drama featuring the same characters but no gender-related content, representing a tighter comparison than control content typically produced for edutainment interventions (Rahmani et al., 2025). Participants were surveyed by phone two weeks before and after the intervention, and a random subsample was interviewed face-to-face six months later by interviewers blinded to respondents' treatment assignment. Additionally, we implement the same experiment in the governorate of Sohag to stress-test treatment effects under different field conditions, including a sample with significantly higher fertility rates and opposition to family planning, and measure outcomes seven

months after the viewing session. Alongside estimating the causal effects of each drama, we develop and test an empirical framework for comparing what is learned from three complementary sources: closed-ended survey responses, interviewer observations, and open-ended interview transcripts analyzed at scale using large language models (LLMS).

We find that the FGM drama bolstered opposition to the practice ($\hat{\beta} = 0.190$ SDs, $p = 0.001$), driven by corrected misperceptions about health risks, fertility, and religious doctrine. Effects of the family planning drama were weak and inconsistent. Both patterns replicate in Sohag despite a more challenging field environment, a more skeptical sample, and outcome measurement seven months post-treatment. Qualitative interviews conducted six months post-treatment corroborate and extend these findings. FGM viewers were 17 percentage points more likely to name FGM as a pressing issue in open-ended responses, and the survey finds increased discussion of the topic among family and friends six weeks later ($\hat{\beta} = 0.360$ SDs, $p < 0.001$). The interviews also reveal depth in the religious misinformation channel: FGM viewers were 28 percentage points more likely to describe religious leaders' opposition to FGM ($p = 0.09$) — the doctrinal correction the state cannot reliably engineer through local clerics. Emotional content analysis shows no elevation of negative affect in the FGM arm, suggesting attitudinal gains do not come at the cost of re-traumatizing participants. The findings indicate that narrative drama can bridge the gap between state and citizen where enforcement and elite co-optation fall short.

# 2 Edutainment, Elites, and Measuring Change

Why should we expect narrative drama to succeed where public policy has faltered? We advance a substantive and methodological contribution in seeking to answer this question. Substantively, prior work identifies social norms and collective expectations as primary drivers of gendered harms, and narrative entertainment as well-positioned to address them by shifting reference points. Building on this work, we propose that edutainment can mitigate gendered

harms through the distinct channel of circumventing unreliable local elites, along with the misconceptions they propagate. Methodologically, we combine a randomized design with blinded qualitative interviews to address how longstanding limitations in how edutainment evaluations are conducted. We use these tools to detect new classes of outcomes, like spontaneous salience, visible distress, and emotional register, to the study of edutainment in ways inform ongoing debates about the ethics of researching sensitive topics.

Narrative entertainment is thought to influence beliefs by encouraging perspective-taking (Cohen et al., 2012; Green and Brock, 2000) and exposing audiences to information, arguments, and social norms without provoking resistance and counter-argument (Petty and Cacioppo, 1986; Slater and Rouner, 2002; Bandura, 2004). This ability to overcome resistance is particularly valuable when messages concern sensitive topics — like FGM or family planning — that socially conservative audiences would otherwise disengage from entirely. The contrast with non-narrative approaches is striking in the Egyptian context specifically. Christia et al. (2023) tested a gender-based violence intervention by an Egyptian women's rights organization delivered via social media and television during COVID-19 and found little evidence of attitude change, with the authors noting that their campaign "differs from edutainment interventions that featured dramatized characters," the gap this study aims to aim to fill, and consistent with the view that the barrier to change is, in part, social.

The evidence that norm-shifting interventions can move behavior at scale in the domain of gendered harms is now considerable, lending natural support to the use of mass media interventions. In the family planning domain, exposure to Brazilian soap operas portraying smaller families led to significant reductions in fertility by changing beliefs and reference points (La Ferrara et al., 2012b). The expansion of cable television in India reduced fertility and improved women's autonomy by reshaping gender norms through entertainment content rather than explicit policy messaging (Jensen and Oster, 2009). Programs addressing early marriage similarly emphasize norms and expectations: combining incentives, information, and empowerment can delay marriage and childbearing in South Asia, particularly when

they alter the perceived returns to girls' education or expand outside options (Buchmann et al., 2022; Bandiera et al., 2020). The experimental evidence on FGM specifically points to marriage-market coordination as the mechanism sustaining the practice. Quasi-experimental variation from anti-FGM radio campaigns in Egypt shows that reducing FGM also lowers bride prices and induces substitution toward other chastity-signaling behaviors, underscoring how deeply economic returns within marriage markets reinforce the norm (Khalifa, 2022). Field-experimental work in Sierra Leone tests norm-replacement strategies directly, offering culturally legitimate alternatives to cutting that preserve the social meaning of initiation rituals while removing the harm (Corno and La Ferrara, 2023).

We propose that gendered harms face a further challenge that this literature has not fully addressed: the problem of unreliable intermediaries. Harmful practices are sticky in gender-conservative societies in part because they are hard to monitor, and because weak state capacity in the peripheries means enforcement is unlikely even where monitoring is possible. The limits of the central state's reach mean that local elites not only shape community norms but even perpetuate the practices themselves, making them impossible to co-opt from above. Experimental evidence demonstrates that trusted authorities can shift norms and reduce harmful beliefs and behaviors (Blair et al., 2021; Siegel and Badaan, 2020). Yet this leverage is futile when those authorities are themselves the problem, exploiting trust in their position to propagate harm (Mattingly, 2016), even using the very same tools of the mass media to do so (Yanagizawa-Drott, 2014). We argue that narrative dramas leapfrog over intermediaries entirely. By reaching rural households directly, edutainment circumvents the need for local elites to transmit the state's message faithfully. This advantage is compounded when elite realms overlap with the nature of misinformation driving these practices. For instance, where practices are driven by underestimating medical risk, then circumventing unreliable doctors — who perpetuate harmful misperceptions as well as practices — is especially powerful.

These substantive arguments motivate equally important methodological choices. Randomized evaluations of narrative entertainment are now common (Rahmani et al., 2025), but

their measurement strategies typically rely on post-treatment surveys that capture whether respondents endorse a position when prompted — not whether that position has become part of how they spontaneously reason about an issue. Qualitative evaluations of edutainment are also widespread, but typically suffer from non-random assignment, immediate post-exposure measurement that inflates demand effects, and researchers who are not blinded to treatment status (cf. Howard-Merrill et al., 2024). Scholars across political science and sociology have long called for embedding qualitative tools within field experiments to capture meaning-making processes that average treatment effects cannot reveal (Paluck and Cialdini, 2014; Paluck, 2010; Small, 2009; Bowers et al., 2024). Our approach responds to these calls by combining randomized exposure with endline interviews conducted six months later by blinded enumerators, analyzed at scale using large language models. This design also generates new evidence on the ethics of researching sensitive topics. The balance of empirical evidence suggests re-traumatization is not a significant concern in social science research: a meta-analysis involving nearly 74,000 participants found that individuals tolerate participation well and frequently report personal benefits such as catharsis and validation (Jaffe et al., 2015), survivors of wartime rape in Eastern Congo reported that speaking to researchers was "relieving," (Aroussi, 2020) and longitudinal research with survivors of intimate partner violence found that participants' fears of re-traumatization did not materialize (Hamberger et al., 2020). By triangulating interviewer-recorded observations of respondent comfort with LLM-coded emotional content of transcripts, we provide a systematic test of whether discussing FGM in a research interview elevates distress relative to a neutral topic.

# 3 Egyptian Context

Egypt has long dominated Arab and African screen culture, with its dozens of home-grown studios producing three-quarters of all feature-length films in the Middle East and North Africa from 1908 to 2007 (Shafik, 2007). The use of Egyptian cinema as an explicit instru-

ment of state policy, however, is more recent. At a military *iftar* dinner during the Ramadan season of 2025, President al-Sisi condemned television dramas that failed to reflect Egyptian values, calling on "artistic works to serve as a mirror to the Egyptian family's deep-rooted values, actively contributing to building a national consciousness worthy of Egypt's status" (Egyptian Presidency, 2026). Prime Minister Madbouly convened a high-level task force to devise a ten-year plan to "activate the role of media and drama in rebuilding the Egyptian personality" and pledged direct state support for "purposeful drama works that achieve social and national goals" (Egypt Independent, 2025; Ahram Online, 2025). Egypt's storied cultural production has also led pundits to characterize its drama industry as a vehicle for regional soft power (Financial Times, 2025), amplifying the stakes of content decisions well beyond domestic audiences.

The practices the state has sought to address through edutainment expose girls and women to serious harm. FGM carries significant medical risks, including obstetric complications, psychological trauma, and mortality (World Health Organization, 2018; Berg and Denison, 2012; Varol et al., 2015). Early marriage and restricted access to family planning increase the likelihood of early, closely spaced pregnancies and constrain women's economic opportunities and household bargaining power. These harms tend to cluster, with FGM frequently framed as a prerequisite for marriageability and both practices sustained by norms regulating girls' sexuality and family honor (UNICEF, 2022). Egypt consistently ranks near the bottom of global gender-parity indices (World Economic Forum, 2025), and regional survey data document widespread support for male breadwinner norms that limit women's labor-market participation, especially outside major cities (Arab Barometer, 2023).

The persistence of these practices points to their deeper social roots. FGM in Egypt, as across much of Africa, predates Islam and is sustained by entrenched norms around purity, explaining its prevalence among both Muslim and Christian communities (Mackie, 1996; World Health Organization, 2018; Blaydes and Platas, 2020). Efforts to address FGM, child marriage, and inadequate family planning must therefore confront not only knowledge gaps

around fertility and health, but also the interlocking norms and belief systems — religious interpretation, marriage-market incentives, and ignorance of health implications — that sustain these practices (Mackie, 1996; Cislaghi and Heise, 2018; Shell-Duncan et al., 2018). For instance, our baseline data shows only 59.9% of respondents in Luxor believed that FGM harms health, a minority (44.2%) believe it has no hadith basis, and 71.3% correctly reject the idea that it improves fertility, demonstrating the intersecting nature of misconceptions (Table 2).

The state has secured the formal cooperation of Al-Azhar and other central religious institutions, but their condemnations of FGM have not consistently reached the village level, where local clerics often maintain that the practice has a religious basis. "I don't care what today's *sheikhs* say. The *sheikhs* of the past were much better, like Sheikh Shaarawi [a state-sponsored cleric] who said that a girl must be cut," one rural Egyptian told researchers (Elnakib et al., 2025). Similarly, medical professionals have now out-paced traditional practitioners in performing FGM, doing so in clinical settings and private homes alike, and operating within local economies of trust and financial incentives that national bans have struggled to penetrate (Elnakib et al., 2025). Parents falsely believe that doctors willing to perform FGM are "very knowledgeable and trustworthy so they will be able to act if anything goes wrong during the procedure," yet only 62% of Egyptian medical students in one study understood that it carried the risk of complications (Mostafa et al., 2006). Physicians who understand these risks are often indifferent to reporting those who do not: "It is no longer my business if they choose to go to someone else. I will not intervene," said one male physician in Minya (Elnakib et al., 2025). Drama sidesteps this problem by packaging belief correction, whether religious or medical, in a narrative that reaches households directly, thereby reducing reliance on local intermediaries.

Within this landscape, Egypt's entertainment industry occupies an outsized role. The country's television sector is among the most developed in the Middle East: virtually all Egyptians (98.8 percent) have access to a working television, the median viewer watches

approximately three hours per day (U.S. Agency for Global Media, 2014), and serialized dramas produced for the Ramadan broadcast season serve as focal points of household viewing and national conversation (Arab Reporters for Investigative Journalism (ARIJ), 2023). The study site of Luxor Governorate is well-suited to examine these dynamics. Subnational data indicate that FGM prevalence exceeds 85%, child marriage is common, and female labor-force participation and educational attainment remain limited (Ministry of Health and Population (Egypt) et al., 2015; Central Agency for Public Mobilization and Statistics, 2021; Ministry of Health and Population et al., 2022), while television ownership and exposure to nationally broadcast programming are universal (U.S. Agency for Global Media, 2014).

Faced with the limits of legal enforcement and public health campaigns in Upper Egypt's rural governorates, the UNFPA launched the 360-Degree Edutainment Campaign with the government's backing in 2022, combining mass-media dissemination with community-based programming to promote adolescent girls' well-being and challenge norms surrounding FGM, child marriage, and family planning. A central component was *Berry Leaves*, a 33-episode, nationally broadcast television drama produced by a major production house and featuring prominent young actors. Script development drew on focus groups and workshops conducted in socioeconomically marginalized communities, with storylines weaving together religious discourse, intergenerational conflict, and everyday moral dilemmas to address harmful practices while remaining culturally legible to audiences in rural areas. We worked with media professionals to identify theme-relevant scenes and merge these as seamlessly as possible into 1.5-hour dramas presenting a coherent storyline.

# 4 Study Design

Partnering with the MENA office of the Abdul Latif Jameel Poverty Action Lab (J-PAL), a research organization housed in the American University of Cairo, we recruited roughly 400

participants in Luxor. This section documents all research procedures for this sample.

We also conduct a parallel study in the more impoverished governorate of Sohag ($n = 595$). A perennial concern in field experiments is that carefully controlled research settings may not generalize to the resource-constrained environments where policy is actually implemented. To probe this, we partnered with UNFPA Egypt and local NGOs to embed an analogous edutainment screening in Sohag. Implementation occurred through community youth centers with limited screening infrastructure, modest incentives constrained by bureaucratic regulations, and considerably less logistical control than the Luxor study (see Appendix A.4 for details). The Sohag sample is also distinct on theoretically meaningful dimensions (Wilke and Humphreys, 2020). Relative to Luxor, Sohag respondents are older, more likely to be married (87% vs. 49%), less educated (9% tertiary vs. 33%), have higher fertility rates (4.1 vs. 3.0 children), and are far more resistant to contraception at baseline (53% vs. 83% pro-contraception after the first birth, (Tables 1 and A4). When it comes to FGM, however, the pattern reverses: only 26% of Sohag respondents believed ex-ante that the hadith supports FGM, compared to 41% in Luxor. This combination of differences allows us to move beyond replication and instead test external validity among populations with different baseline attitudes toward the behaviors under study.

## 4.1 Recruitment and Sampling

Sampling relied on convenience-based recruitment, with enumerators going door-to-door seeking adult residents — who had not seen the show before — to invite to the edutainment study. In total, 399 individuals (315 women and 84 men) provided informed consent. Of these, 336 had schedules that were compatible with being assigned to a viewing session, and were therefore eligible for randomization. One week after recruitment, the baseline survey was administered to all consented participants, and two weeks after the baseline, the viewing sessions began. Participants were assigned to viewing groups stratified by gender and by participants' preferred day and time to attend the session. The resultant viewing sessions,

or 'clusters', were then allocated to one of three arms. Overall, 81 participants were assigned to the placebo group, 133 to the early marriage and FGM group, and 122 to the family planning group. Balance across the treatment arms is shown in Table A2, and descriptive statistics are shown along demographic (Table 1) and attitudinal dimensions (Table 2).

Table 1: Descriptive Statistics by Treatment Arm — Luxor

| | Placebo ($N = 79$) | FGM Reel ($N = 132$) | FP Reel ($N = 121$) | Overall ($N = 332$) |
|---|---|---|---|---|
| *Demographics* | | | | |
| Age (mean, SD) | 31.1 (10.8) | 32.8 (10.7) | 33.5 (10.5) | 32.6 (10.6) |
| Female (%) | 48.1 | 95.5 | 91.7 | 82.8 |
| *Marital status* | | | | |
| Married (%) | 43.0 | 43.9 | 59.5 | 49.4 |
| Single (%) | 53.2 | 43.2 | 31.4 | 41.3 |
| Widowed / divorced (%) | 3.8 | 12.9 | 9.1 | 9.3 |
| *Education* | | | | |
| No formal / primary (%) | 10.1 | 16.7 | 17.4 | 15.4 |
| Secondary (%) | 45.6 | 52.3 | 55.4 | 51.8 |
| Tertiary or above (%) | 44.3 | 31.1 | 27.3 | 32.8 |
| *Household* | | | | |
| Avg. no. children (among parents) | 3.1 (1.4) | 2.9 (1.2) | 3.1 (1.3) | 3.0 (1.3) |
| Has working TV (%) | 97.5 | 97.0 | 97.5 | 97.3 |
| *Attitudes* | | | | |
| FGM religiously required/encouraged (%)† | 32.9 | 24.2 | 31.4 | 28.9 |
| Pro-contraception (%)‡ | 77.2 | 87.9 | 82.6 | 83.4 |

*Notes:* Statistics computed from baseline data (full sample). Age and number of children reported as mean (SD); all other entries are percentages. Education binned from a 12-category scale. Single includes never-married and engaged respondents; widowed/divorced includes separated. The female share varies by arm because randomization was stratified by gender.

† Share responding that FGM is "religiously required" or "religiously encouraged" (*baseline*).

‡ Share agreeing or strongly agreeing: "It is appropriate for a couple to use family planning after the first birth" (*baseline*).

## 4.2 Treatments

Each viewing session was randomly assigned to one of the following groups, all of which watched a 1.5 hour highlight drama covering the following themes:

### 4.2.1 Early Marriage and FGM

This drama follows Dalila, a young woman forced into marriage before the age of 18 by her grandmother, a practitioner of the Zar folk ritual, practiced in lower Egypt and believed to

Table 2: Baseline Attitudes by Domain and Treatment Arm

| | Placebo | FGM Reel | FP Reel | Overall |
|---|---|---|---|---|
| *FGM & Early Marriage* | | | | |
| Opposes FGM continuation | 74.4 | 71.3 | 64.0 | 69.5 |
| FGM harms health | 63.4 | 64.8 | 52.3 | 59.9 |
| FGM unrelated to fertility | 72.1 | 70.0 | 72.3 | 71.3 |
| FGM has no basis in hadith | 45.1 | 50.0 | 37.1 | 44.2 |
| Supports legal minimum marriage age | 69.6 | 74.8 | 72.0 | 72.6 |
| *Family Planning* | | | | |
| Supports FP after first birth | 77.2 | 89.9 | 84.7 | 85.0 |
| FP improves family finances | 76.6 | 82.3 | 76.7 | 78.9 |
| Equal say in FP decisions | 92.2 | 95.4 | 97.5 | 95.4 |
| Ideal no. of children (mean, SD) | 2.8 (1.2) | 2.6 (0.7) | 2.9 (0.8) | 2.8 (0.9) |
| *Other Gender Attitudes* | | | | |
| Supports women working | 61.0 | 45.2 | 46.4 | 49.5 |
| Rejects wife-beating | 26.6 | 11.4 | 20.8 | 18.4 |
| Opposes father choosing husband | 72.7 | 63.0 | 71.8 | 68.5 |
| Equal say in children's future | 100.0 | 96.2 | 98.3 | 97.9 |

*Notes:* Entries are the percentage agreeing or strongly agreeing among respondents with an opinion (excluding "don't know" responses), except ideal number of children which is reported as mean (SD). All variables measured at baseline.

expel evil spirits and ease emotional suffering. Dalila's husband dies shortly after, leaving her a pregnant minor. Because the father had passed away and the mother was underage, the grandmother falsely registered the baby, named Anhar, as Dalila's sister rather than her daughter. As her daughter Anhar grows up, the grandmother pressures Dalila to have the girl circumcised. Dalila, scarred by her own traumatic circumcision, resists and seeks guidance from scholars at Al-Azhar, where she learns that FGM has no basis in religious doctrine and poses serious health risks. Her grandmother overrules her objections. Anhar is circumcised, develops an untreated infection, and is forced into marriage at fifteen. Her health deteriorates through pregnancy and complications from the procedure; she is eventually hospitalized, but dies. The drama conveys its central messages — that FGM is medically harmful and religiously unjustified, and that early marriage compounds these harms — through Anhar's fate and Dalila's journey from victimhood to resistance.

### 4.2.2 Family Planning

The family planning drama follows two parallel storylines. In the first, Hassan is the eldest son of a large, impoverished family of eight crowded into a single room. Hassan works as a janitor in an acting school, and is envious of the wealthier students. His burdens are compounded by the financial irresponsibility of his father, who squanders the family's earnings at local cafes. His mother, already overwhelmed, discovers she is pregnant again. Her daughter persuades her to consult a doctor, who recommends an IUD as a long-term contraceptive option, while offering birth control pills and condoms as well, to which the mother confirms that her husband would never use a condom. When Hassan's sister is sexually assaulted on a cleaning job, her father chooses to accept hush money from the assailant. Hassan is frustrated by his father's decision, as well as familial pressures to spend his salary not on bettering his own situation and marrying the girl he loves, but instead to save for his sister's wedding.

The second story centers on Marian, a Christian woman whose husband, desperate for a son after two daughters, pressures her into repeated pregnancies. She suffers multiple miscarriages, falls into depression, and her health deteriorates until her family finally accepts medical advice that another pregnancy would be life-threatening. Together, the storylines portray large families as a source of financial strain, physical suffering, and constrained opportunity, making the case for family planning in recognizable domestic circumstances rather than abstract health messaging.

### 4.2.3 Placebo

The placebo drama was taken from the same TV drama and focuses on the same characters, but uses storylines that do not not involve any gender-related topics. The drama follows Sooko, a terminally ill young man known for his kindness toward peers at the acting school where he works as a janitor. After his mother dies, Sooko is injured in a confrontation. Neighbors rush him to the hospital where he receives treatment. Despite the violence, Sooko

chooses not to report the incident or seek revenge. Sooko is later found dead, the cause ambiguous between his illness, the injury, and grief. The drama holds format and dramatic tone constant while varying content.

# 5 Empirical Strategy

We investigate the effects of random assignment to each drama on gender-related attitudes, norms, and emotional responses. Because our estimates condition on post-randomization attendance at the viewing session, they are best understood as Complier Average Causal Effects (CACEs), or the average effect of watching the assigned drama among participants who showed up. Each viewing session was randomly assigned to one of three dramas, making the session the unit of randomization. Standard errors are clustered accordingly. We focus on results comparing each treatment group to the pooled control (the other drama plus placebo), which yields the highest precision. We present placebo-only comparisons as a secondary analysis.

Outcomes are measured in two complementary ways: (1) a phone-based survey administered two weeks after the viewing sessions, and (2) semi-structured, in-depth interviews conducted six months later with a randomly selected subsample. The survey questionnaire and interview guide were designed in parallel to measure the same core constructs, allowing us to evaluate not only whether edutainment shifts standard survey outcomes, but also whether it alters narrative framing, emotional engagement, and conversational ease around sensitive topics. The experimental analysis using survey-based outcomes are pre-registered with the Open Science Foundation (OSF) for Luxor (Z9e4y) and Sohag (G56v8), while LLM-based outcomes are pre-registered separately (OSF 64dzq). Both studies have received IRB approval (UCLA protocol 24-0455).

## 5.1 Survey Outcomes

A phone survey conducted two weeks after the viewing sessions achieved 98.8% retention: 332 of 336 randomized participants completed the endline. Enumerators were centralized in a single call center to ensure supervision and quality control; participants received 50 EGP per completed survey and 100 EGP plus refreshments for attending the viewing session.

We group primary survey outcomes into five thematic domains: FGM attitudes, family planning preferences, and early marriage (core outcomes reflecting drama content), plus women's empowerment and responses to domestic violence as secondary outcomes to explore spillover effects (see Table A8 for full item lists). For both treatments, we measure two classes of mechanisms: an information pathway (beliefs about health risks and religious justifications for FGM, knowledge of contraceptive options) and a social norms pathway (perceived community support, discussing the behavior with family and friends). For FGM, we also test an empathy pathway (perspective-taking toward affected women). This last pathway extends the edutainment literature, which has focused more on message content than on whether empathy with characters itself drives change (Rahmani et al., 2025), despite a growing evidence base on parasocial contact (Alrababah et al., 2021; Schiappa et al., 2005; Lotun et al., 2022).

Within each outcome family, we construct indices to increase precision. For our primary composite measures — the FGM Opposition and FGM Health Information indices, each comprising three items — we fit a structural equation model (SEM) and extract factor scores, allowing item weights to reflect each indicator's contribution to the underlying latent construct. All remaining indices consist of two items and are constructed as the row mean of their standardized components. All indices and individual outcomes are then standardized to the mean and standard deviation of the placebo group.

## 5.2 In-Depth Interviews and Interviewer Ratings

To capture more nuanced responses to the dramas than what closed-form surveys might capture, we conducted face-to-face qualitative interviews with a random sample of sixty participants — blocked by treatment assignment and randomly selected from each cluster, with interviewers blinded to treatment status.[2] With semi-structured questions covering FGM, family planning, early marriage, women's education and employment, and domestic violence, interviews averaged 45–60 minutes and were transcribed from the original Arabic. Forty-nine of the sixty respondents completed the interview (82%), an attrition rate that was stable across treatment arms.[3] Each session concluded with a brief post-interview survey in which the interviewer reported the respondent's comfort, perceived openness or reluctance to discuss each topic, and stress. The latter is measured through a prespecified list of observable indicators, like fidgeting, whispering, or avoiding eye contact, rather than subjective impressions.

From the transcripts and interviewer reports, we construct four sets of outcomes. First, we measure *topic salience*: whether respondents spontaneously name FGM, early marriage, or family planning as a pressing issue facing women in their community. We do so using LLMs, which are well-suited to coding the variety of colloquial Egyptian Arabic expressions respondents use to describe these issues. Second, *survey-parallel outcomes* are produced by coding open-ended responses to interview questions that we designed to mirror questions taken from the closed-form survey, allowing us to directly compare the same construct measured in interview versus survey format (see Appendix C.1 for the full mapping). We code the interview transcripts via a structured LLM prompt aligned with our survey measures.[4]

[2]Only twice did respondents reveal their treatment status during the interview, confirming the integrity of the blinded design.

[3]We observe attrition rates of 20% for the FGM and family planning dramas, and 15% for FGM.

[4]We use two LLMs for coding—GPT-4o and GPT-5.1—which differ in their decision-making procedures. To validate this approach, we manually review and revise the models' outputs for all variables. The median agreement rate between the initial LLM coding and the human-revised coding is 97% for GPT-4o and 93%

Third, *interviewer ratings*, responses to a brief checklist of behavioral indicators filled out by the interviewer immediately after each session, capture constructs that self-reports cannot easily assess — whether respondents appeared at ease, reluctant, or responding to social pressure on sensitive topics. Fourth, we measure the *emotional content* of transcripts using an LLM coding protocol applied to respondents' verbatim Arabic answers for each topic area (FGM, family planning, early marriage). The LLM is given a fixed emotion label set (Table 3) and asked to return up to five emotion labels ranked by prevalence, accompanied by short Arabic snippets from the transcript justifying each assignment. To assess reliability, we run the procedure using two LLM models (GPT-4o and GPT-5.1). The two models show high overall consistency, with an average agreement rate of approximately 85% across emotion categories.[5] This measure allows us to assess whether discussing sensitive topics like FGM elevates emotional distress among treated respondents relative to the placebo group.

for GPT-5.1, with median Cohen's kappa values of 0.90 and 0.78, respectively. For the salience measures, agreement rates range from 94% to 100%. Appendix C.6 reports these validation results for all variables. Throughout the analysis, we use the human-revised versions of the LLM-coded outcomes.

[5]Appendix C.6.2 presents additional statistics on the alignment between the two models' sentiment analyses. We use GPT-5.1 for the main analysis and relegate GPT-4o replications to the appendix.

| Emotion | Arabic example | English translation |
|---|---|---|
| Resignation / acceptance | احنا طلعنا لقيناها كده. ما نقدرش نستغنى عنها عشان احنا طلعنا لقيناها كده دي عاداتنا | "We grew up finding it this way. We cannot give it up, because we were raised with it like this; these are our customs." |
| Disgust / aversion | مالوش مميزات، كله قرف | "It has no advantages; it's all disgusting." |
| Sadness / grief | دي تجربة قاسية ومؤلمة جدًا... أنا عن نفسي اتعذبت من ده جدًا. | "It is a very harsh and painful experience... I personally suffered from it a great deal." |
| Anger / indignation | ظلم... والله ده ظلم بيّن للبنت. | "It is — truly, this is clear injustice against the girl." |
| Fear / anxiety | في بعض البنات ممكن يحصل لهم مضاعفات أثناء الختان، زي النزيف... لو البنت اتأذت منه، بيتحول لمصيبة. | "Some girls can suffer complications during circumcision, such as bleeding... if the girl is harmed by it, it becomes a disaster." |
| Compassion / empathy | لأن البنوتة بتتأثر نفسيًا جدًا، وتفقد ثقتها في نفسها، ومبتنساش اللحظة دي طول حياتها. | "Because the little girl is affected psychologically very deeply, loses confidence in herself, and never forgets that moment for the rest of her life." |
| Hope / optimism (change) | آه الجيل اللي طالع وعيه مختلف... لا للختان. | "Yes, the rising generation is different in its awareness... no to circumcision." |
| Pride / dignity | الست اللي بتشتغل بتكون منفتحة أكتر، وواثقة من نفسها، ومعاها فلوسها اللي بتحسسها بالأمان. | "A woman who works becomes more open-minded, more confident in herself, and has her own money, which gives her a sense of security." |
| Relief / comfort | الشغل أحسن جدًا... ومعاها فلوسها اللي بتحسسها بالأمان. | "Work is much better... and having her own money makes her feel secure." |
| Confusion / uncertainty | انا نفسي متلخبطة... الواحد مبقاش عارف ايه الصح و ايه لأ. | "I myself am confused... a person no longer knows what is right and what is not." |
| Neutral / flat | ايوه منتشرة... عادات وتقاليد قديمة. | "Yes, it is widespread... old customs and traditions." |

Table 3: Example interview snippets illustrating the full LLM-coded emotion categories.

# 6 Results

## 6.1 Quantitative Outcomes

We start by estimating treatment effects on outcomes directly related to the themes of each narrative. Starting with the FGM drama, the first panel of Figure 1 shows that exposure

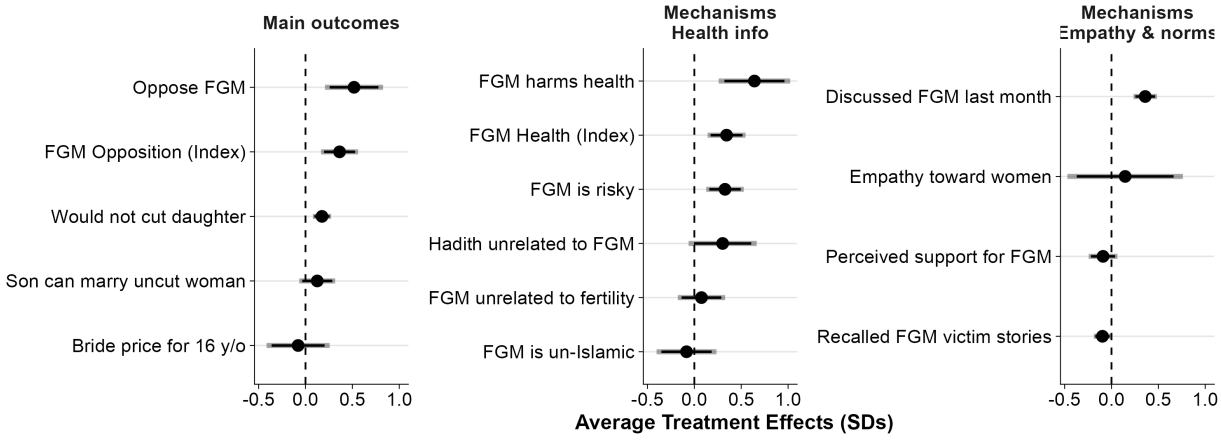



Figure 1: Average Treatment Effects: FGM Drama

**Notes:** Positive values indicate increases relative to the placebo group and Family Planning group, in standardized units. Coefficients based on an OLS regression controlling for gender, age, education, marital status, and the specific baseline item where available (see text). The FGM Health Info Index and FGM Opposition Index are constructed using SEMs to match the measurement model approach used in the companion analysis. Cronbach's $\alpha = 0.67$ and 0.77, respectively.

to this drama had significant effects on attitudes and behavioral intentions most closely tied to the storyline (Table A5). The first row of the figure indicates a significant increase in opposition to FGM as a practice ($\hat{\beta} = 0.518$ SDs, $p = 0.008$), while the third row shows a significant increase in refusal to circumcise one's daughters ($\hat{\beta} = 0.176$ SDs, $p = 0.005$); a gain of 17.6 percentage points relative to a placebo-group mean of 65.3% who were willing to make this commitment. Effects on the composite FGM Opposition Index, which uses structural equation modeling to estimate treatment effects on a latent factor comprising opposition to FGM, willingness to allow one's son to marry an uncircumcised woman, and intention not to circumcise future daughters, are also large ($\hat{\beta} = 0.190$ SDs, $p = 0.001$). More equivocal is the effect on allowing one's son to marry an uncircumcised woman and the effect on dowries for child brides, which did not shift upward or toward "no dowry is acceptable" ($\hat{\beta} = 0.000$, $p = 0.997$).[6]

[6]This survey item is coded as an ordinal variable, with "no dowry is acceptable" reflecting the top-coded responses, and all others reflecting ordered price levels in descending order from the equivalent of USD 125,000 to USD 200, according to 2025 exchange rates.

Turning to the outcomes that shed light on causal mechanisms (Figure 1), the drama strengthened the belief that FGM harms women's health ($\hat{\beta} = 0.641$ SDs, $p = 0.008$) relative to a baseline of 59.9% (Table 2), and that it is generally risky as opposed to safe ($\hat{\beta} = 0.327$ SDs, $p = 0.009$), a finding the drama dramatizes in poignant fashion. Effects on the composite FGM Health Information Index are also large, positive, and precisely estimated ($\hat{\beta} = 0.326$ SDs, $p = 0.003$). Effects on the religious and Hadith-related items are smaller, although respondents do seem to update their beliefs on the lack of Hadith-based and religious-based evidence for FGM in the correct direction (Hadith: $\hat{\beta} = 0.163$, $p = 0.136$; evidence for FGM in Islam: $\hat{\beta} = -0.083$, $p = 0.620$). We note that 27% of respondents did not know how to answer the *hadith* item at baseline, an uncertainty that helps contextualize movement in this domain. These results suggest that the drama's primary informational pathway runs through health, but that the dramas also sowed some doubt about the religious basis for FGM.

The right-hand panel illustrating shifts in empathy and social norms presents a more ambiguous picture. Relative to the placebo group, treated respondents were 36 percentage points more likely to report having discussed FGM with "relatives, friends, or neighbors" in the previous month ($\hat{\beta} = 0.360$ SDs, $p < 0.001$), relative to a placebo-group mean of 27.8%. However, effects on empathy toward women and perceived community support for FGM are essentially null. Notably, the likelihood of recalling "any stories of people suffering health complications as a result of the FGM procedure" is somewhat negative (recalled FGM victim stories: $\hat{\beta} = -0.096$, $p = 0.066$), although respondents may have interpreted this question as pertaining to personal acquaintances rather than characters in the drama.

Figure 2 and Table A6 present analogous results for the family planning drama. The main outcomes panel offers mixed evidence of effectiveness. The ideal number of children moves in the expected direction but falls short of significance after controlling for baseline preferences ($\hat{\beta} = 0.115$ SDs, $p = 0.237$; approximately 0.12 fewer children relative to the placebo group mean of 2.67). Effects on the intent to advise one's adult children to use family planning

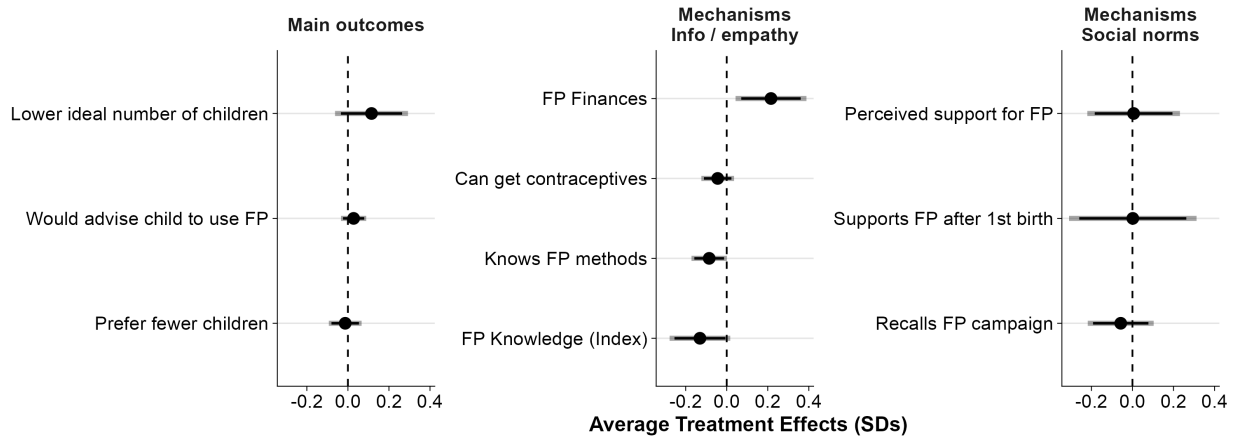



Figure 2: Average Treatment Effects: Family Planning Drama

**Notes:** Positive values indicate increases relative to the placebo group and FGM group, in standardized units. Coefficients based on an OLS regression controlling for gender, age, education, marital status, and the baseline ideal number of children for FP main outcomes and knowledge items, and baseline contraceptive use for FP norms items.

methods and on the preference for fewer children are near zero ($\hat{\beta} = 0.028$, $p = 0.400$ and $\hat{\beta} = -0.013$, $p = 0.753$, respectively). The mechanisms panels reinforce this murky picture: the FP Knowledge Index is marginally negative ($\hat{\beta} = -0.130$, $p = 0.118$), driven by the effect on knowledge of specific FP methods ($\hat{\beta} = -0.085$, $p = 0.085$). FP drama viewers were also no more likely to report that they know where to get contraceptives, despite the free clinics portrayed in the show. The one mechanism that does move in the anticipated direction is the belief that family planning relieves the burden on family finances ($\hat{\beta} = 0.216$ SDs, $p = 0.038$), which was a central theme in the dramatic narrative. Effects on social norms like perceived community support for family planning, the appropriateness of a couple to use family planning after the first birth, and recalling family planning messages from radio, television, or print media in the last six months are all near zero. This is consistent with the near-ceiling baseline values on family planning attitudes (Table 2), which leave limited scope for the drama to move.

As a robustness check, we replicate all analyses using the placebo-only control group in place of the pooled control (Figures A1 and A2 in Appendix B.2; Tables A9 and A10). For the FGM drama, the two specifications yield broadly consistent results. Opposition to

FGM is large and significant under both comparisons ($\hat{\beta} = 0.518$, $p = 0.008$ and $\hat{\beta} = 0.583$, $p = 0.016$), as is the FGM Opposition Index (pooled: $\hat{\beta} = 0.190$, $p = 0.001$; placebo-only: $\hat{\beta} = 0.228$, $p = 0.005$) and the intention not to circumcise one's daughters (pooled: $\hat{\beta} = 0.176$, $p = 0.005$; placebo-only: $\hat{\beta} = 0.187$, $p = 0.024$). For the family planning drama, all estimates are near zero and insignificant under either specification, reinforcing the conclusion that the FP drama's impacts are weaker and less consistent than those of the FGM drama. The absence of FP drama effects cannot be attributed to weak treatment delivery: correct recall of the assigned drama was high across all arms — 95.5% in the FGM arm, 86.0% in the FP arm, and 94.9% in the placebo (Table A1).

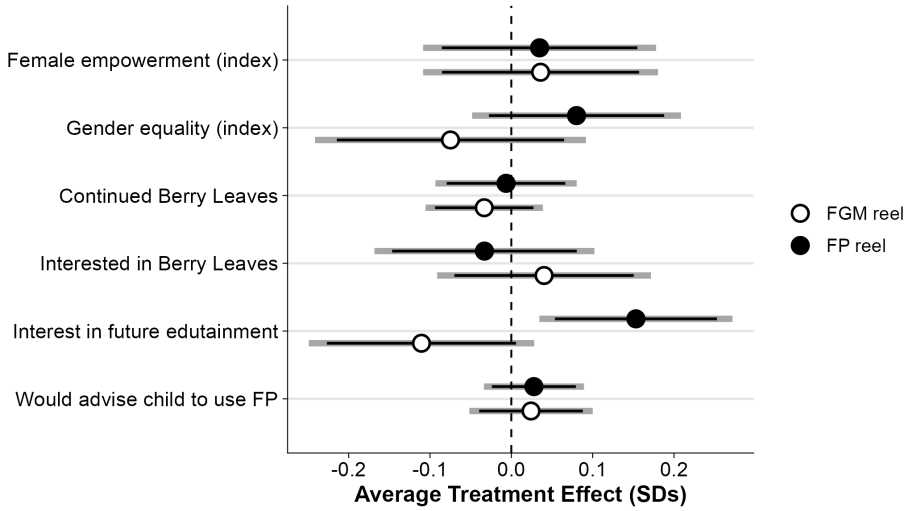



Figure 3: Drama Effects on Broader Gender Attitudes and Media Engagement

**Notes:** Open circles denote FGM drama estimates; filled circles denote FP drama estimates. "Female empowerment" consists of 3 items: whether boys' education should be prioritized over girls', whether fathers should choose their daughters' husbands, and whether women should work. "Gender equality" consists of two items: whether husbands and wives should have equal say in their children's future, and regarding family planning decisions. Positive values indicate increases relative to the placebo group, in standardized units. Coefficients based on an OLS regression controlling for gender, age, education, marital status, and a baseline question on attitudes toward women working. Cronbach's $\alpha = 0.45$ and 0.46, respectively.

Did either drama succeed in changing broader attitudes and behaviors? Effects on secondary outcomes are largely null for both dramas. Gender indices capturing female em-

powerment and gender equality, interest in continuing to watch *Berry Leaves*, and media engagement all hover near zero for both treatment arms (Figure 3; Table A7).[7] The same null pattern holds across all gender-based violence outcomes, like rejection of wife-beating, attitudes toward women reporting violence, and community intervention norms, for both dramas (Figure A3 in Appendix B.3). Only 18.4% of respondents reject wife-beating at baseline (Table 2), indicating a mismatch between the magnitudes of the treatment and social problem it aims to mitigate. The only deviation from the broader pattern of null results is that the FP drama produces a significant increase in interest in future edutainment programming ($\hat{\beta} = 0.153$ SDs, $p = 0.033$), while the FGM drama perhaps has the reverse effect ($\hat{\beta} = -0.110$, $p = 0.149$). These results are consistent with the view that edutainment-induced attitude change is narrow, affecting outcomes directly dramatized in the narrative but not spilling over into adjacent domains.

## 6.2 External Validity in Sohag

The FGM drama produced more robust and policy-relevant effects than the family planning drama, operating through health, and to a lesser extent, religious information rather than empathy or social norms. Do these results travel to different communities? The robustness of the FGM-over-FP differential is corroborated by the study in Sohag. The results are shown in Figures 4 and 5. There, we likewise observe null family planning effects regardless of how we define the control group, while FGM point estimates remain positive despite noisier implementation conditions and a sample more opposed to FGM at baseline. That the same pattern of results holds seven months after the viewing sessions concluded in Sohag suggests that our results represent a lower bound on the sort of change achievable under typical implementation conditions.

[7] About 78% of respondents reported being interested in the show, although this includes 42% who did not have access to viewing means, which helps explain why only 12% reported watching the show on their own.

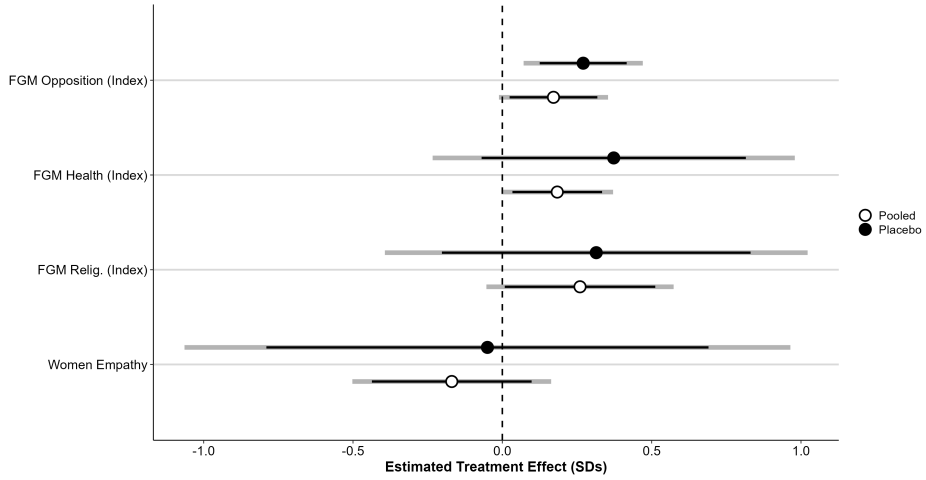



Figure 4: Effect of FGM Drama — Sohag

**Notes:** Coefficients based on an OLS regression controlling for gender, age, education, marital status, viewing date, village, and the baseline variable where possible. The FGM Health Attitudes Index consists of three items: whether FGM harms women's health, is risky, and the misconception that FGM improves fertility. The FGM Opposition Index consists of three items: support for one's son marrying an uncut woman, opposition to circumcising one's future daughters, and opposition to FGM overall. Circles denote estimates using the pooled control group; triangles denote placebo-only. Tabulated results are in A23.

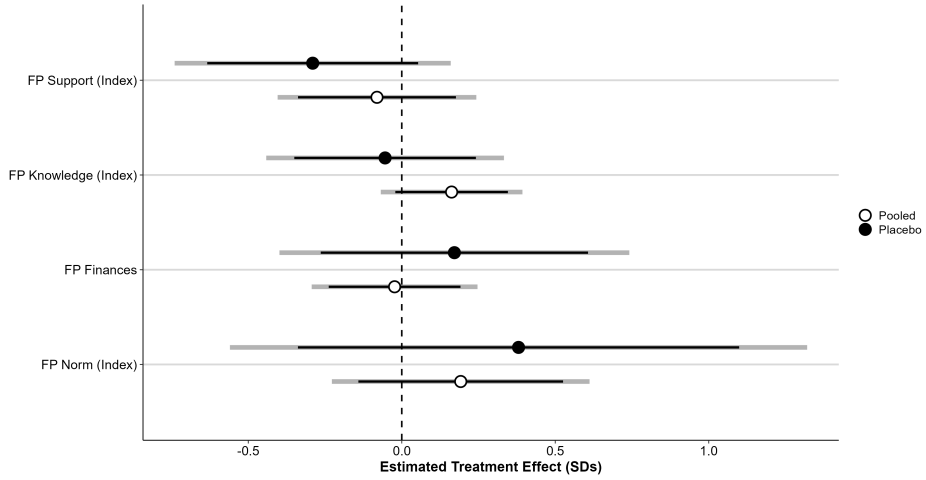



Figure 5: Effect of FP Drama — Sohag

**Notes:** Coefficients based on an OLS regression controlling for gender, age, education, marital status, viewing date, village, and the baseline variable where possible. Circles denote estimates using the pooled control group; triangles denote placebo-only. Tabulated results are in A24.

Additionally, we observe similar results for the subset of respondents who were subsequently interviewed face-to-face (see Figures A5, A6, and A7 in the Appendix). We next turn to those results.

## 6.3 Qualitative Outcomes

*Topic salience.* Six months after treatment, blinded interviewers asked respondents what important issues women face in their community, a question designed to elicit spontaneous salience of the drama topics. The responses reveal a striking contrast between the two dramas. Approximately one-third of those in the FGM treatment group spontaneously mentioned FGM, compared to roughly one-fifth of those in the placebo or pooled control group. Family planning, by contrast, was rarely top-of-mind; mentioned by only 12 percent of FP-treated respondents and none in either control group.

*Survey-parallel outcomes.* Coded interview responses offer a broadly positive picture of FGM attitude change, and correlate moderately with their corresponding midline survey outcomes ($\rho = 0.33$–$0.58$, Table 4). Support for laws banning FGM is substantially higher among the treated (71%) than in the placebo (41%) or pooled control (44%), and an LLM-generated index of overall opposition to FGM is significantly higher among the treated relative to the pooled control.[8] Effects on willingness to allow sons to marry uncircumcised women and intentions not to circumcise daughters are positive but modest. The most vivid illustration comes from one interviewee: "My cousin was engaged, and [the family] had wanted to circumcise her …After I watched the film, I told them in detail about what happened, and they abandoned the idea.'' She added: "Some of my husband's sisters were with us and watched the film, and honestly most people who watched it changed their minds.''

Analysis of the informational mechanisms adds nuance to our survey findings by showing that medical and religious information have different half-lives. Six months after the

[8]Table A13 lists the variables, the estimated differences in means, and the confidence interval on this difference.

treatment, health risks were no more salient among the treated relative to either control group. However, correction of religious misconceptions, a shift only weakly detected in our survey results, persists and gains further prominence: 65% of FGM-treated viewers reject religious justifications for FGM compared to 47% in the placebo and 37% in the pooled control. A significantly higher share of the treated (47%) perceive religious leaders' opposition to FGM, relative to 19% among control groups, and treated respondents are more likely to hold that the Qur'an offers no support for the practice and that proponents rely on unreliable *hadiths*. The interviews shed light on why these shifts are hard-won. Many respondents noted that religious figures "do not discuss FGM in the villages," that "sermons are rarely dedicated to it," and that "some religious figures avoid the topic to accommodate the social context." Others cited doctrinal disagreements between state-affiliated and local clerics: "state-affiliated religious figures oppose it in line with the state, but less famous (local) ones say it has religious bases."

Consistent with the survey results, we see no evidence of treatment effects on empathy toward women or perceived community support for FGM. The interviews nonetheless illuminate how social norms sustain the practice. The marriage-market link appears weaker than often assumed: only 57% of interviewees report that FGM improves a woman's marriage prospects, and 70% refute the idea that the practice is sustained by marriage negotiations (Mackie, 1996). Where marriage-market pressures do operate, respondents trace them to elderly women in the family — mothers, grandmothers, in-laws — rather than to grooms, suggesting that interventions should target women as the primary gatekeepers of this norm.

Shifting to family planning, interviews corroborate the null results in the survey data (Table A14). Support for family planning is high and uniform across all conditions — more than 90% of respondents would advise children to use family planning methods — while salience of the topic remains low across all arms. The sharpest treatment-linked difference involves attitudes toward government regulation, whereby the treated group is far less likely to support state rewards or penalties to promote smaller families (33% vs. 81% in the placebo).

Most opponents frame family size as a personal economic decision, preferring awareness-raising over mandates. Where the interviews do reveal differences by treatment arm is in how respondents reason about financial motivations, already shown in the survey data to be the family planning outcome most amenable to change. Treated respondents were more likely to share personal and second-hand experiences of financial burdens from large families (69%) relative to the placebo (31%) and pooled control (42%). This finding highlights a dimension of persuasion that standard survey measures may understate — respondents may internalize a message by connecting it to lived experience, rather than updating their beliefs in the abstract.

Table 4: Correlations Between Qualitative and Midline Survey Measures

| | $\rho$(Qual–Midline) |
|---|---|
| **FGM Attitudes** | |
| Would Approve Son Marrying Uncut Woman | 0.327 |
| Would Advise Against Circumcising Daughters | 0.577 |
| Oppose FGM Continuation | 0.500 |
| Presence of Religious Evidence for FGM | 0.157 |
| FGM Religiously Allowed | 0.446 |
| Support Laws Banning FGM | 0.375 |
| Aware of FGM Health Harms | 0.350 |
| Empathy/Recollection of FGM Victims | 0.439 |
| **Family Planning** | |
| Advising Children to Use FP Methods | −0.043 |
| Salience of Financial Motivations for FP | 0.050 |
| Contraceptive Use is Appropriate (Norms) | 0.058 |
| Support for Gov. Actions to Limit Family Size | 0.299 |
| Believe in Ideal Number of Children (<4) | 0.246 |

*Notes:* Pearson correlations between qualitative interview-derived measures and corresponding midline survey outcomes.

*Interviewer ratings.* Interviewer-assessed measures of respondent comfort, measured by perceived comfort (1–4 scale), number of observable stress signs, and openness versus reluctance to discuss FGM, are statistically indistinguishable across treatment and placebo groups

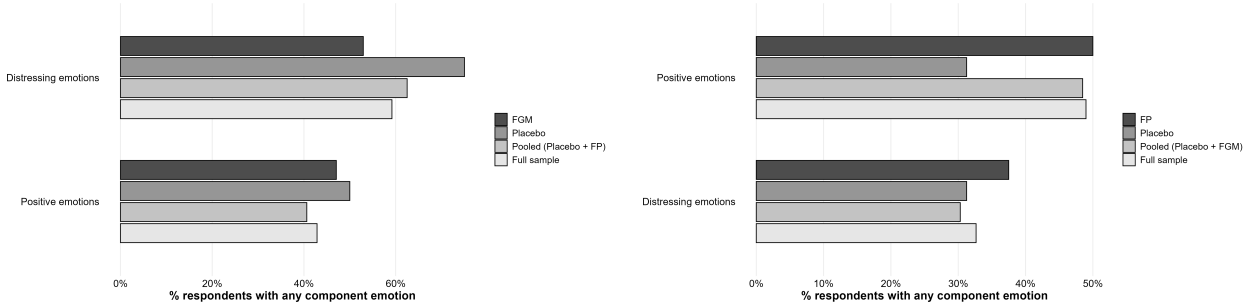
on all three dimensions, and, if anything, display marginally greater openness when FGM is raised (Figure A4 in Appendix C.2). This behavioral evidence reinforces the transcript-based finding that sensitive topics do not translate into observable distress.

*Emotional content.* Figure 6 summarizes the emotional content of interview transcripts for each treatment arm. Positive and distressing emotional responses are broadly balanced across both topic areas: anger, fear, and anxiety are prevalent in FGM discussions but coupled with hopeful, compassionate, and neutral language, and a similar balance holds for family planning. Crucially for interpretation, negative emotions reflect moral motivation rather than personal distress. One interviewee's anger translated into a demand for accountability: "Of course they [perpetrators] should be punished, because this causes death and other illnesses.'' Fear emerged from awareness of consequences: "it negatively affects the child, especially if there is a problem during the (FGM) operation.'' Sadness inspired calls for prohibition: "lots of people lost their daughters because of FGM, I hope it is banned because the harm is evident.''

Most importantly for concerns about re-traumatization, the FGM arm shows no elevation in negative emotional content relative to placebo. Figure 6 contrasts the distribution of distressing emotions with that of positive ones, broken down by treatment arm. Anger and indignation are equally prevalent across arms (47% FGM vs. 50% placebo), fear and anxiety are *lower* in the FGM arm (35% vs. 50%), and sadness and grief stand at 12% versus 25% in the placebo — a gap of 13 percentage points running opposite to what re-traumatization would predict. In family planning discussions, positive emotions prevail, especially relative to the placebo.[9] Sensitive topics track with some emotional engagement, helping to bolster construct validity, without inducing distress beyond what we observe in the control groups.[10]

[9]Full comparisons of emotion prevalence across groups are reported in Appendix C.5, including cross-topic difference-in-differences estimates.

[10]These conclusions remain robust in replications using GPT-4o (C.6.3) and after considering the intensity (not solely the presence) of different emotions using their rank-weighted prevalence (C.6.4).




(a) FGM (b) Family Planning

Figure 6: Aggregate Emotional Content in Interview Transcripts

**Notes:** Share of respondents for whom distressing (anger, fear, sadness, shame, disgust) versus positive (compassion, hope, pride, relief) emotions appear among the five most salient LLM-coded emotions. Panel (a): FGM question set; panel (b): family planning question set. Percentages may exceed 100% because multiple emotions may be coded per respondent.

# 7 Discussion

*Berry Leaves* exemplifies a pattern increasingly scrutinized by political scientists: states using narrative entertainment not just to suppress dissent (Esberg, 2020) but to actively cultivate regime-aligned beliefs and compliance with central directives (Mattingly and Yao, 2022; Yao, 2024; Deng, 2025). What distinguishes the Egyptian case is that the state turned to drama as a response to the limits of more direct approaches. Legislation, enforcement, and official public health campaigns had not stopped moving the needle in Upper Egypt's rural governorates, where FGM and child marriage are sustained by entrenched norms and local authorities, both religious and medical, that the center cannot fully co-opt. Narrative dramas tackle this problem by delivering doctrinal and scientific correction directly to households. Our results suggest that edutainment is most effective when it can circumvent unreliable intermediaries in this way, when the practice it targets is sustained by correctable misperceptions, and when both the practice and its dramatic depiction carry emotional weight.

The FGM drama satisfies all three conditions while the family planning drama satisfies

none. FGM is viscerally disturbing and generates sustained conversation long after viewing. Family planning lacks that emotional charge and does not suffer from the same legitimacy gap between the state and its local agents. Preferences around family size are likely shaped more by economic policies, as pointed out by one interviewee: "Nowadays, the government made it so that [food ration cards] cover the mother, father, and two children. So the state helped push in this direction too." We find that two distinct forms of misperceptions, scientific beliefs grounded in objective facts and doctrinal ones reflecting subjective religious interpretations, were amenable to change. This second pathway echoes experimental evidence from Egypt showing that Qur'anic reinterpretations in favor of women's political leadership shifted support for it more effectively than a secular argument alone, because opposition to women's leadership was itself grounded in religion (Masoud et al., 2016).

Somewhat paradoxically, the emotional resonance of the FGM drama and its subsequent staying power in participants' minds reflect the same dynamic that animates ethical concerns about researching sensitive topics. When we analyze the emotional content of interview transcripts using LLMs six months after the intervention, however, we find no elevation of fear, anger, or grief among FGM-treated respondents relative to the placebo group. Anger, if anything, is more prevalent among placebo respondents when discussing FGM; sadness and grief are rare across all conditions. The topic is salient and emotionally laden, but that emotional charge takes the form of moral indignation and hope rather than personal distress. The family planning drama, despite enjoying the same level of production value, could not overcome the fact that contraception and family size remain socially unremarkable relative to FGM. These results add to a growing body of evidence that the risks of re-traumatization in research on sensitive topics may have been overstated (Weiss, 2023; Jaffe et al., 2015), and highlight that the instinct to avoid such topics in edutainment may itself carry a cost, forgoing the emotional engagement that makes attitude change possible.

Equally notable are the mechanisms that failed to move: social norms. At baseline, 74% of respondents personally supported a legal ban on FGM and 73% supported banning child

marriage, positions that align with state policy. Yet respondents simultaneously estimated that 57% of their community supports FGM and 54% supports early marriage. The null results we uncover on social norms for either drama suggests that the gap between personal opposition and perceived social norms is sticky. These results are consistent with the broader edutainment literature, which finds that social norms are considerably more resistant to narrative persuasion than individual attitudes (Rahmani et al., 2025) — and may help explain why prevalence remains high even as stated opposition rises.

# 8 Conclusions

Our results identify one reason why state policy has has struggled to eradicate gendered harms, and point toward a potential remedy. When a state bans a practice that communities experience as culturally and religiously grounded, enforcement depends, in part, on convincing those in the peripheries that the practice is not just illegal but also "wrong" in the terms that govern rural life, backed by trusted authorities in the church, mosque, or clinic. In Upper Egypt, televised dramas embed corrective information — both scientific and doctrinal — in a way that circumvents the need for local clerics to cooperate, and emotionally resonates with audiences in a way that gives the drama's core messages staying power months after the show concludes.

These nuanced patterns would have been difficult to detect with any single data source. The study develops and tests a framework for integrating blinded qualitative interviews into experimental evaluations of harmful practices, combining pre-registered survey outcomes, LLM-coded interview transcripts, and interviewer-recorded behavioral observations. The combination allows us to estimate causal effects with the precision of a randomized design, capture emotional and salience dimensions invisible to survey instruments, and assess whether the research process itself harms participants. On the last question, we reassuringly find that sensitive topics do not elevate distress relative to neutral ones, and the properties

that make them sensitive — emotional resonance and persistent salience — are also what make them attractive targets for attitude change. Standard evaluations relying solely on post-treatment surveys may therefore understate the degree to which edutainment embeds issues into everyday conversation (Paluck, 2009; Papa et al., 2000). Future work should deploy both instruments at similar time intervals to determine whether the qualitative channel detects persistence that surveys eventually lose, or whether it captures dimensions of change that ultimately do not surface in closed-ended responses.

We note that our study sites of Luxor and Sohag share the same national media environment, and, crucially, the same backdrop of explicit government endorsement of anti-FGM and pro-family planning messaging. One might worry that our results are inflated by the Egyptian state's official endorsement of the drama's messages. If state authority were sufficient, however, there would have been little need for edutainment in the first place. As one interviewee put it: "The law tried to limit it, but it's still there." Nevertheless, the more stringent tests — whether edutainment can achieve comparable effects in the absence of state legitimation, where religious and medical authorities are ambivalent rather than officially opposed, and where laws forbidding harmful practices have not yet taken shape — are ones that can define the next generation of research on edutainment, public policy, and behavioral change.

# References

Ahram Online (2025). Egypt forms committee to draft roadmap for media reform.

Al Arabiya English (2017). Egypt imam accused of marrying off underage girls. Accessed January 2026.

Alrababah, A., Marble, W., Mousa, S., and Siegel, A. A. (2021). Can exposure to celebrities reduce prejudice? the effect of mohamed salah on islamophobic behaviors and attitudes. *American Political Science Review*, 115(4):1111–1128.

Arab Barometer (2023). Gender gaps in mena remain stubbornly entrenched despite positive trends. Summarizes traditional gender-attitude patterns in Egypt and the MENA region.

Arab Gulf States Institute in Washington (2022). Ramadan series tackle women's issues and religion. Accessed January 2026.

Arab Reporters for Investigative Journalism (ARIJ) (2023). Egyptian drama suffers from a shrinking creative space and freedoms. Includes reporting on annual volume of Egyptian TV drama production. Accessed January 2026.

Aroussi, S. (2020). Researching wartime rape in eastern congo: why we should continue to talk to survivors? *Qualitative Research*, 20(5):582–597.

Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., and Sulaiman, M. (2020). Women's empowerment in action: Evidence from a randomized control trial in africa. *American Economic Journal: Applied Economics*, 12(1):210–259.

Bandura, A. (2004). Social cognitive theory for social and personal change enabled by the media. In Singhal, A., Cody, M., Rogers, E. M., and Sabido, M., editors, *Entertainment-Education and Social Change: History, Research, and Practice*, pages 75–95. Lawrence Erlbaum Associates, Mahwah, N.J.

Barsoum, G., Crépon, B., Dyer, P., Gardiner, D., Michel, B., Moaz, M., and Parienté, W. (2022). Evaluating the impact of entrepreneurship edutainment in Egypt: An experimental approach. *Economica*, 89(353):82–109.

Berg, R. C. and Denison, E. (2012). Does female genital mutilation/cutting (fgm/c) affect women's sexual functioning? a systematic review of the sexual consequences of fgm/c. *Sexuality Research and Social Policy*, 9:41–56.

Blair, G., Littman, R., Nugent, E. R., Wolfe, R., Bukar, M., Crisman, B., Etim, A., Hazlett,

C., and Kim, J. (2021). Trusted authorities can change minds and shift norms during conflict. *Proceedings of the National Academy of Sciences*, 118(42):e2105570118.

Blaydes, L. and Platas, M. R. (2020). Religion, family structure, and the perpetuation of female genital cutting in egypt. *Journal of Demographic Economics*, 86(3):305–328.

Bowers, J., Green, D. P., and Aronow, P. M. (2024). Designing and analyzing mixed-methods experiments. *PLOS ONE*, 19(2):e0324936.

Buchmann, N., Field, E., Glennerster, R., Nazneen, S., Pimkina, S., and Sen, M. (2022). Power vs. money: Alternative approaches to reducing child marriage in bangladesh. *Journal of Political Economy*. Online early / working paper versions circulated from 2018.

Central Agency for Public Mobilization and Statistics (2021). Luxor governorate statistical yearbook.

Christia, F., Larreguy, H., Parker-Magyar, E., and Quintero, M. (2023). Empowering women facing gender-based violence amid COVID-19 through media campaigns. *Nature Human Behaviour*, 7:1740–1752.

Cislaghi, B. and Heise, L. (2018). Four avenues of norm change: A conceptual framework for social norms interventions. *Global Public Health*, 13(11):1774–1782.

Cohen, J., Shavitt, S., and Solomon, S. (2012). Exploring subtext processing in narrative persuasion. *Journal of Communication*, 62(4):566–589.

Corno, L. and La Ferrara, E. (2023). Norms replacement and information: A field experiment on ending female genital cutting. *Journal of Development Economics*, 162.

Daily News Egypt (2022). Al-azhar condemns the Faten Amal Harby series and warns of marginalising the sunnah. Accessed January 2026.

Deng, R. W. (2025). Screened realities: How entertainment fosters political compliance in autocracies. Revise and Resubmit, *American Political Science Review*.

Egypt Independent (2025). PM says gov't ready for supporting 'purposeful' media, drama.

Egyptian Presidency (2026). Presidential statement on drama and national values.

El-Dabbee, R., Fahmy, A., Youssry, D., and El-Gibaly, O. (2022). Religion, authority, and female genital mutilation/cutting in egypt: Perspectives from religious leaders and health providers. *Reproductive Health Matters*, 30(1):1–12. Documents mixed and conflicting religious views regarding FGM/C among Egyptian clerics.

Elnakib, S., Sallab, M., Hussein, A., Marouf, M., Elshiwy, S., Elsherbini, R., Daggula, K., and Prihartono, I. (2025). Medicalization of female genital mutilation in egypt: Trends, drivers, and prospects for elimination. *PLOS Global Public Health*, 5(11):e0005195.

Esberg, J. (2020). Censorship as reward: Evidence from pop culture censorship in Chile. *American Political Science Review*, 114(3).

Financial Times (2025). Egypt's soft power and the drama industry.

Green, M. C. and Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5):701–721.

Hamberger, L. K., Larsen, S., and Ambuel, B. (2020). It helped a lot to go over it: Intimate partner violence research risks and benefits from participating in an 18-month longitudinal study. *Journal of Family Violence*, 35(1):43–52.

Howard-Merrill, L., Pichon, M., Witt, A., Sono, R., Gimunta, V., Hofer, E., Kiluvia, F., and Buller, A. M. (2024). "I Have Learned That Nothing Is Given for Free": An exploratory qualitative evaluation of a social norms edutainment intervention broadcast on local radio to prevent age-disparate transactional sex in kigoma, tanzania. *BMC Public Health*, 24:2907.

Jaffe, A. E., DiLillo, D., Hoffman, L., Haikalis, M., and Dykstra, R. E. (2015). Does it hurt

to ask? a meta-analysis of participant reactions to trauma research. *Clinical Psychology Review*, 40:40–56.

Jensen, R. and Oster, E. (2009). The power of tv: Cable television and women's status in india. *Quarterly Journal of Economics*, 124(3):1057–1094.

Khalifa, S. (2022). Female genital cutting and bride price. Job Market Paper, Aix–Marseille School of Economics.

La Ferrara, E., Chong, A., and Duryea, S. (2012a). Soap operas and fertility: Evidence from brazil. *American Economic Journal: Applied Economics*, 4(4):1–31.

La Ferrara, E., Chong, A., and Duryea, S. (2012b). Soap operas and fertility: Evidence from brazil. *American Economic Journal: Applied Economics*, 4(4):1–31.

Lapinski, M. K. and Nwulu, P. (2008). Can a Short Film Impact HIV-Related Risk and Stigma Perceptions? Results from an Experiment in Abuja, Nigeria. *Health Communication*, 23(5):403–412.

Lauby, J. L., LaPollo, A. B., Herbst, J. H., and Painter, T. M. (2010). Preventing AIDS through Live Movement and Sound: Efficacy of a Theater-Based HIV Prevention Intervention Delivered to High-Risk Male Adolescents in Juvenile Justice Settings. *AIDS education and prevention*, 22(5):402–416.

Lotun, S., Sankaran, S., Vivekananda-Schmidt, P., Bowles, D., Novarese, M., Dolan, R. J., and Fleming, S. M. (2022). Parasocial relationships on youtube reduce prejudice towards mental health issues. *Scientific Reports*, 12(1):18401.

Mackie, G. (1996). Ending footbinding and infibulation: A convention account. *American Sociological Review*, 61(6):999–1017.

Masoud, T., Jamal, A., and Nugent, E. (2016). Using the qur'ān to empower arab women?

theory and experimental evidence from egypt. *Comparative Political Studies*, 49(12):1555–1598.

Mattingly, D. C. (2016). Elite capture: How decentralization and informal institutions weaken property rights in china. *World Politics*, 68(3):383–412.

Mattingly, D. C. and Yao, E. (2022). How soft propaganda persuades. *Comparative Political Studies*, 55(9).

Michaelson, R. (2023). Hit TV drama sparks calls for reform of egypt's oppressive guardianship law. Accessed January 2026.

Ministry of Health and Population, El-Zanaty and Associates, and ICF (2022). Egypt family health survey 2021: Key indicators report.

Ministry of Health and Population (Egypt), El-Zanaty and Associates, and ICF International (2015). Egypt demographic and health survey 2014.

Mohamed, A. E. (2024). The relationship between state-provided Islamic education and islamism. *Politics and Religion*, 17(3):468–493.

Mostafa, S., El Zeiny, N., Tayel, S., and Moubarak, E. (2006). What do medical students in alexandria know about female genital mutilation? *EMHJ-Eastern Mediterranean Health Journal, 12 (Supp. 2), S78-S92, 2006.*

Moustafa, T. (2000). Conflict and cooperation between the state and religious institutions in contemporary Egypt. *International Journal of Middle East Studies*, 32(1):3–22.

Paluck, E. L. (2009). Reducing intergroup prejudice and conflict using the media: A field experiment in rwanda. *Journal of Personality and Social Psychology*, 96(3):574–587.

Paluck, E. L. (2010). The promising integration of qualitative methods and field experiments. *The ANNALS of the American Academy of Political and Social Science*, 628(1):59–71.

Paluck argues that integrating qualitative and quantitative methods allows exploration of mechanisms and that qualitative work can be central to experimental designs.

Paluck, E. L. and Cialdini, R. B. (2014). Field research methods. In Reis, H. T. and Judd, C. M., editors, *Handbook of Research Methods in Social and Personality Psychology*, pages 81–104. Cambridge University Press, Cambridge, 2 edition.

Paluck, E. L., Porat, R., Clark, C. S., and Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual review of psychology*, 72(1):533–560.

Papa, M. J., Singhal, A., Law, S., Pant, S., Sood, S., Rogers, E. M., and Shefner-Rogers, C. L. (2000). Entertainment-education and social change: An analysis of parasocial interaction, social learning, collective efficacy, and paradoxical communication. *Journal of Communication*, 50(4):31–55.

Petty, R. and Cacioppo, J. (1986). *Communication and Persuasion: Central and Peripheral Routes to Attitude Change. New York: Springer-Verlag*. Springer-Verlag, New York.

Rahmani, B., Montano, B., Groves, D. W., and Green, D. P. (2025). The persuasive effects of narrative entertainment: a meta-analysis of recent experiments. *Behavioural Public Policy*, page 1–23.

Schiappa, E., Gregg, P. B., and Hewes, D. E. (2005). The parasocial contact hypothesis. *Communication Monographs*, 72(1):92–115.

Shafik, V. (2007). *Popular Egyptian cinema: Gender, class, and nation*. Oxford University Press.

Shell-Duncan, B., Naik, R., and Feldman-Jacobs, C. (2018). A state-of-the-art synthesis on female genital mutilation/cutting: What do we know now? *Evidence to End FGM/C Program Report*.

Siegel, A. A. and Badaan, V. (2020). no2sectarianism: Experimental approaches to reducing sectarian hate speech online. *American Political Science Review*, 114(3):837–855.

Slater, M. D. and Rouner, D. (2002). Entertainment-Education and Elaboration Likelihood: Understanding the Processing of Narrative Persuasion. *Communication Theory*, 12(2):173–191.

Small, M. L. (2009). 'how many cases do i need?' on science and the logic of case selection in field-based research. *American Journal of Sociology*, 114(1):5–38.

South, E. N. (2024). Breaking the silence: How egypt's fight against female genital mutilation is transforming lives. Accessed January 2026.

The New Arab (2022). Egyptian lawyer accuses TV series on divorce of blasphemy. Accessed January 2026.

UNFPA Arab States Regional Office (2020). Child marriage in egypt: A statistical profile. Reports 7.1 million women married before 18 and 1.3 million before 15.

UNICEF (2016). Female genital mutilation/cutting: A global concern. Includes Egypt DHS 2015 prevalence estimate of 87.5% among women aged 15–49.

UNICEF (2022). Understanding the relationship between child marriage and female genital mutilation/cutting. Global survey analysis showing patterns of co-occurrence in many countries.

Unplugged, R. (2023). Ramadan TV dramas as a conduit for social and legal change. Accessed January 2026.

U.S. Agency for Global Media (2014). Egypt: Media use and public opinion. Reports 98.8% household TV ownership and high daily television use. Accessed January 2026.

Varol, N., Fraser, I. S., Ng, C. H., Jaldesa, G., and Hall, J. (2015). Female genital mutilation/cutting – towards abandonment of a harmful practice. *Australian and New Zealand Journal of Obstetrics and Gynaecology*, 55:400–405.

Weiss, A. (2023). Beyond retraumatization: Trauma-informed political science research.

Wilke, A. and Humphreys, M. (2020). Field experiments, theory, and external validity. *L. Curini & R. Franzese, The SAGE Handbook of Research Methods in Political Science and International Relations*, pages 1007–1035.

World Economic Forum (2025). Global gender gap report 2025.

World Health Organization (2018). Care of girls and women living with female genital mutilation: A clinical handbook.

Yanagizawa-Drott, D. (2014). Propaganda and conflict: Evidence from the rwandan genocide. *The Quarterly Journal of Economics*, 129(4):1947–1994.

Yao, L. (2024). Popular entertainment as propaganda: Insights from an online field experiment in China. Working paper, Vanderbilt University.

# Table of Contents

**A Pre-Registration and Study Design**

A.1 Deviations from Pre-Analysis Plan 1

A.2 Manipulation Check 1

A.3 Balance Across Treatment Arms 2

A.4 Sohag Study 2

**B Experimental Results**

B.1 Regression Tables 5

B.2 Robustness Checks 7

B.3 Secondary Outcomes: Gender-Based Violence 9

B.4 Sohag Results 26

**C Qualitative Measurement**

C.1 Survey–Interview Alignment 11

C.2 Interviewer-Coded Outcomes 12

C.3 Qualitative Subsample Results 12

C.4 Survey-Aligned Qualitative Outcomes 16

C.5 Emotional Reactions Across Groups 19

C.6 LLM Coding Validation 21

**D Research Instruments**

D.1 Interview Guide **??**

D.2 Post-Interview Report Form **??**

D.3 Survey–Interview Question Mapping **??**

# A Pre-Registration and Study Design

## A.1 Deviations from Pre-Analysis Plan

This study is governed by three pre-analysis plans: one covering the survey-based outcomes for the Luxor study (OSF: Z9e4y), the Sohag study (G56v8), and the post-intervention qualitative interviews (64dzq). We note four deviations. First, the interview PAP pre-specifies interviewer-rated measures of respondent comfort and stress (`comfort_pint`, `stress_pint`) but does not anticipate the LLM-based emotional coding reported in Section 6.3. The 10-label emotion taxonomy, dual-run stability check, and re-traumatization analysis are therefore exploratory additions, motivated by the ethical questions raised during data collection. Second, both PAPs pre-specify heterogeneous treatment effects by gender, education, and baseline gender attitudes; we do not report these analyses due to power constraints. Third, the FGM Opposition Index replaces the pre-specified `rate_fgm_risksafe` item with `fgm_son_marriage_ml`; the substitution was made to improve internal consistency (Cronbach's $\alpha = 0.77$ with the substituted item vs. a lower value with the original). Fourth, the survey outcome capturing whether or not the respondent has discussed FGM in the past month was not listed in any of the PAPs. This was an oversight.

## A.2 Manipulation Check

Table A1: Manipulation Check: Correct Recall of Assigned Drama

| | FGM Drama | FP Drama | Placebo |
|---|---|---|---|
| Total $N$ | 132 | 121 | 79 |
| Recalled correct story ($N$) | 126 | 104 | 75 |
| **% correct recall** | **95.5%** | **86.0%** | **94.9%** |

*Notes:* Respondents were asked to describe the episode they had just watched. The "correct" response is the option with the highest share in each arm: the FGM-focused narrative for FGM-arm respondents, the family-planning narrative for FP-arm respondents, and the placebo (*Berry Leaves*) narrative for placebo-arm respondents.

## A.3 Balance Across Treatment Arms

Table A2: Baseline balance across treatment arms

| Variable | Mean (SD) [N] Placebo | FGM | FP | Std. diff. FGM vs Placebo | FP vs Placebo |
|---|---|---|---|---|---|
| *Demographics* | | | | | |
| Age | 31.1 (10.8) [79] | 32.8 (10.7) [132] | 33.5 (10.5) [121] | 0.152 | 0.224 |
| Married | 0.43 (0.50) [79] | 0.44 (0.50) [132] | 0.60 (0.49) [121] | 0.018 | 0.330 |
| Number of children | 0.05 (1.14) [32] | -0.12 (0.93) [67] | 0.08 (1.00) [79] | -0.148 | 0.031 |
| Education (bucket) | 0.19 (1.05) [79] | 0.02 (0.97) [132] | -0.14 (0.98) [121] | -0.161 | -0.311 |
| *Family planning (baseline)* | | | | | |
| Contraceptive use | -0.29 (1.16) [79] | 0.11 (0.86) [132] | 0.07 (1.00) [121] | 0.341 | 0.304 |
| Equal say in family planning | -0.19 (1.15) [79] | 0.05 (0.99) [131] | 0.07 (0.90) [121] | 0.207 | 0.227 |
| Household finances for FP | -0.06 (0.99) [79] | 0.09 (0.97) [132] | -0.06 (1.03) [121] | 0.149 | -0.006 |
| *FGM attitudes (baseline)* | | | | | |
| Support for continuation | 0.09 (0.96) [79] | 0.04 (1.00) [132] | -0.10 (1.02) [121] | -0.059 | -0.206 |
| Beliefs about fertility | 0.06 (0.92) [78] | -0.03 (1.04) [131] | -0.00 (1.01) [121] | -0.101 | -0.066 |
| Perceived health harms | 0.09 (0.94) [78] | 0.11 (0.99) [131] | -0.18 (1.03) [121] | 0.017 | -0.285 |
| *Gender norms (baseline)* | | | | | |
| Women working outside the home | 0.21 (0.97) [79] | -0.09 (1.01) [132] | -0.04 (0.99) [121] | -0.314 | -0.259 |

*Notes:* For each baseline variable, we report the mean, standard deviation, and number of non-missing observations separately for the Placebo, FGM, and FP arms. Standardized differences are computed relative to the Placebo group as $(\bar{x}_{\text{treat}} - \bar{x}_{\text{placebo}})/sd_{\text{placebo}}$, using only non-missing observations within each arm.

## A.4 Sohag Study

The Sohag and Minya replication study was conducted in partnership with UNFPA Egypt, the youth peer network Y-Peer/EFPA, the civil society organization Etijah, and the Egyptian Ministry of Youth, which provided access to youth center facilities and program rosters. The study targeted governorates with among the highest fertility rates in Egypt according to the Demographic and Health Survey: Sohag, Minya, Assiut, and Beni Suef. Qualitative data collection focused on two villages in Sohag governorate, Soflaq and al-Sawallem, where 50 semi-structured interviews were conducted with household members of study participants.

The experimental design mirrored the Luxor study: eligible participants who had not watched *Berry Leaves* were randomly assigned to view the same FGM/early marriage drama, family planning drama, or placebo drama in community youth centers, stratified by village and gender. The pilot phase launched in late 2023; door-to-door recruitment was conducted

in mid-2024 after phone-based outreach proved insufficient. The post-intervention survey was administered 7 months after the viewing sessions. Of the 536 participants who attended and completed the follow-up survey, 234 were assigned to the FGM drama, 145 to the FP drama, and 157 to the placebo, organized across 22 viewing clusters. Table A3 summarizes the core study features.

Table A3: Sohag Study: Overview

| **Feature** | **Details** |
|---|---|
| Locations | Sohag governorates; qualitative fieldwork in Soflaq and al-Sawallem (Sohag) |
| Implementing partners | UNFPA Egypt, Y-Peer/EFPA, Etijah, Ministry of Youth |
| Timing | Pilot: late 2023; door-to-door recruitment: mid-2024; post-intervention survey: 7 months post-screening |
| Treatment arms | FGM/early marriage drama, family planning drama, placebo (same content as Luxor) |
| Randomization | Cluster-based, stratified by village and gender; 22 clusters |
| Analytic sample | $N = 536$ (FGM: 234; FP: 145; Placebo: 157) |
| Qualitative | 50 semi-structured interviews with participants' household members |

*Notes:* Outreach attempted to reach 11,000 individuals by phone; 3,508 completed the baseline survey; 536 attended viewing sessions and completed the follow-up.

### A.4.1 Implementation Challenges

Several challenges make the Sohag and Minya study a more conservative test than Luxor. Initial recruitment relied on outdated NGO contact lists, necessitating a door-to-door campaign; despite enrolling over 3,400 participants at baseline, logistical and transport constraints reduced the analytic sample to 536 (roughly 84% attrition). Screening quality varied across youth centers and NGO facilities, with some sessions using small screens and low audio. Partner organizations differed in staffing and community relationships, introducing noise into cross-session comparisons. Targeted villages had prior exposure to UNFPA-affiliated campaigns, potentially compressing measurable change. Finally, outcomes were measured 7 months post-screening — a longer lag than Luxor's two weeks. Any effects surviving this environment likely understate the true causal impact; the null FP results and attenuated FGM estimates should be interpreted accordingly.

Table A4: Descriptive Statistics by Treatment Arm — Sohag

| | Placebo ($N = 153$) | FGM Reel ($N = 228$) | FP Reel ($N = 144$) | Overall ($N = 525$) |
|---|---|---|---|---|
| *Demographics* | | | | |
| Age (mean, SD) | 39.0 (13.1) | 40.6 (12.8) | 38.6 (12.1) | 39.6 (12.7) |
| Female (%) | 92.8 | 79.4 | 89.6 | 86.1 |
| *Marital status* | | | | |
| Married (%) | 87.6 | 86.0 | 87.5 | 86.9 |
| Single (%) | 3.9 | 3.1 | 5.6 | 4.0 |
| Widowed / divorced (%) | 8.5 | 11.0 | 6.9 | 9.1 |
| *Education* | | | | |
| No formal / primary (%) | 39.9 | 38.2 | 34.7 | 37.7 |
| Secondary (%) | 56.2 | 53.5 | 50.0 | 53.3 |
| Tertiary or above (%) | 3.9 | 8.3 | 15.3 | 9.0 |
| *Household* | | | | |
| Avg. no. children (among parents) | 4.0 (1.7) | 4.2 (1.7) | 4.0 (1.7) | 4.1 (1.7) |
| Has working TV (%) | 96.1 | 96.5 | 96.5 | 96.4 |
| *Attitudes* | | | | |
| FGM religiously required/encouraged (%)† | | *pooled control: 34.7* | | — |
| Pro-contraception (%)‡ | 57.5 | 50.4 | 53.5 | 53.3 |

*Notes:* Statistics computed from baseline data (full sample). Age and number of children reported as mean (SD); all other entries are percentages. Education binned from a 12-category scale. Single includes never-married and engaged respondents; widowed/divorced includes separated. The female share varies by arm because randomization was stratified by gender.

† Share responding that FGM is "religiously required" or "religiously encouraged." *Midline only* in Sohag; figure shown is for the pooled control group (placebo + FP arm, $N = 297$). By-arm breakdown not shown to avoid contamination.

‡ Share agreeing or strongly agreeing: "It is appropriate for a couple to use family planning after the first birth" (*baseline*).

# B Experimental Results

## B.1 Regression Tables

Table A5: Average Treatment Effects: FGM Drama

| Outcome | $\hat{\beta}$ (SDs) | SE | $p$-value | |
|---|---|---|---|---|
| *Panel A: Main outcomes* | | | | |
| **Oppose FGM** | **0.518** | **0.181** | **0.008** | *** |
| **FGM Opposition Index** | **0.190** | **0.058** | **0.001** | *** |
| **Would not cut daughter** | **0.176** | **0.057** | **0.005** | *** |
| Son can marry uncut woman | 0.125 | 0.102 | 0.227 | |
| Bride price attitudes (index) | 0.000 | 0.083 | 0.997 | |
| *Panel B: Mechanisms — Health information* | | | | |
| **FGM harms health** | **0.641** | **0.221** | **0.008** | *** |
| **FGM is risky** | **0.327** | **0.117** | **0.009** | *** |
| **FGM Health Info Index** | **0.326** | **0.108** | **0.003** | *** |
| FGM unrelated to fertility | 0.076 | 0.133 | 0.569 | |
| Hadith unrelated to FGM | 0.163 | 0.100 | 0.136 | |
| FGM is un-Islamic | −0.083 | 0.167 | 0.620 | |
| *Panel C: Mechanisms — Empathy & social norms* | | | | |
| **Recalled FGM victim stories** | **−0.096** | **0.050** | **0.066** | * |
| Empathy toward women | 0.146 | 0.388 | 0.705 | |
| Perceived support for FGM | −0.090 | 0.082 | 0.276 | |
| **Discussed FGM last month** | **0.360** | **0.085** | **<0.001** | *** |

Positive values in SDs of placebo-group distribution. OLS with CR2 cluster-robust SEs. Controls: gender, age, education, marital status, plus the specific baseline item listed in the text. SEMs used to create FGM Opposition and Health Info Indices. **Bold** = $p \leq 0.10$. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A6: Average Treatment Effects: Family Planning Drama

| Outcome | $\hat{\beta}$ (SDs) | SE | $p$-value | |
|---|---|---|---|---|
| *Panel A: Main outcomes* | | | | |
| Lower ideal number of children | 0.115 | 0.096 | 0.237 | |
| Prefer fewer children | −0.013 | 0.041 | 0.753 | |
| Would advise child to use FP | 0.028 | 0.033 | 0.400 | |
| *Panel B: Mechanisms — Information & finances* | | | | |
| FP Knowledge Index | −0.130 | 0.081 | 0.118 | |
| Can get contraceptives | −0.044 | 0.043 | 0.307 | |
| **Knows FP methods** | **−0.085** | **0.048** | **0.085** | * |
| **FP improves finances** | **0.216** | **0.098** | **0.038** | ** |
| *Panel C: Mechanisms — Social norms* | | | | |
| Recalls FP campaign | −0.057 | 0.085 | 0.504 | |
| Supports FP after 1st birth | 0.002 | 0.188 | 0.992 | |
| Perceived support for FP | 0.006 | 0.123 | 0.962 | |

See Table A5 notes. Baseline control: ideal number of children (baseline) for FP main outcomes and knowledge items; contraceptive use (baseline) for FP norms. **Bold** = $p \leq 0.10$. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A7: Drama Effects on Broader Gender Attitudes and Media Engagement

| | FGM drama | | FP drama | | |
|---|---|---|---|---|---|
| Outcome | $\hat{\beta}$ | $p$ | $\hat{\beta}$ | $p$ | |
| Female empowerment (index) | 0.036 | 0.636 | 0.035 | 0.646 | |
| Gender equality (index) | −0.075 | 0.399 | 0.080 | 0.253 | |
| Continued Berry Leaves | −0.033 | 0.386 | −0.006 | 0.887 | |
| Interested in Berry Leaves | 0.040 | 0.562 | −0.033 | 0.644 | |
| **Interest in future edutainment** | −0.110 | 0.149 | **0.153** | **0.033** | ** |
| Would advise child to use FP | 0.024 | 0.546 | 0.028 | 0.400 | |

**Bold** = $p \leq 0.10$. Estimates in SDs relative to placebo group. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A8: Constructed Indices with Cronbach's $\alpha > 0.40$

| Index | $\alpha$ | Items | Components |
|---|---|---|---|
| FGM Opposition | 0.777 | 3 | Opposition to continuation of FGM; willingness for son to marry uncut woman; intention not to circumcise daughter |
| FGM Health Info | 0.670 | 3 | Belief that FGM harms health; belief FGM does not improve fertility; perception that FGM is risky |
| FP Knowledge | 0.417 | 2 | Knowledge of access to contraceptives; awareness of multiple family planning methods |
| Female Empowerment | 0.452 | 3 | Support equal education for girls; oppose father choosing husband; support women working |
| Gender Equality | 0.456 | 2 | Equal say in family planning; equal say over daughter's future |

## B.2 Robustness Checks

Figures A1 and A2 compare main outcome and mechanism estimates using the pooled control group (placebo plus second treatment arm) against the placebo-only control. Circles denote pooled-control estimates; triangles denote placebo-only estimates. Tables A9 and A10 report the corresponding regression output.

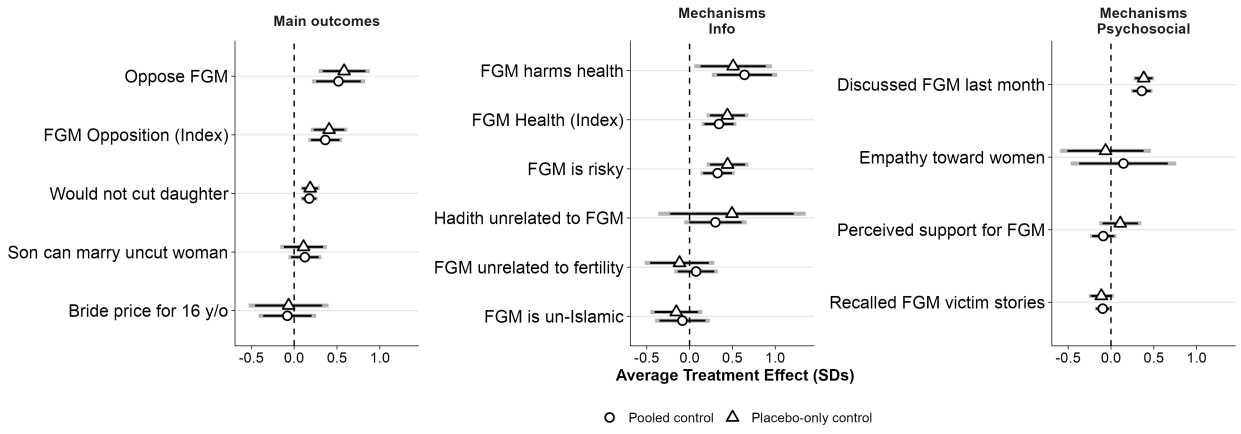



Figure A1: Average Treatment Effects: FGM Drama (Robustness)

**Notes:** Circles denote estimates using the pooled control group (placebo plus second treatment arm); triangles denote estimates using the placebo-only control group. Positive values indicate increases relative to the control group, in standardized units. Coefficients based on an OLS regression controlling for gender, age, education, marital status, and the baseline variable where possible. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

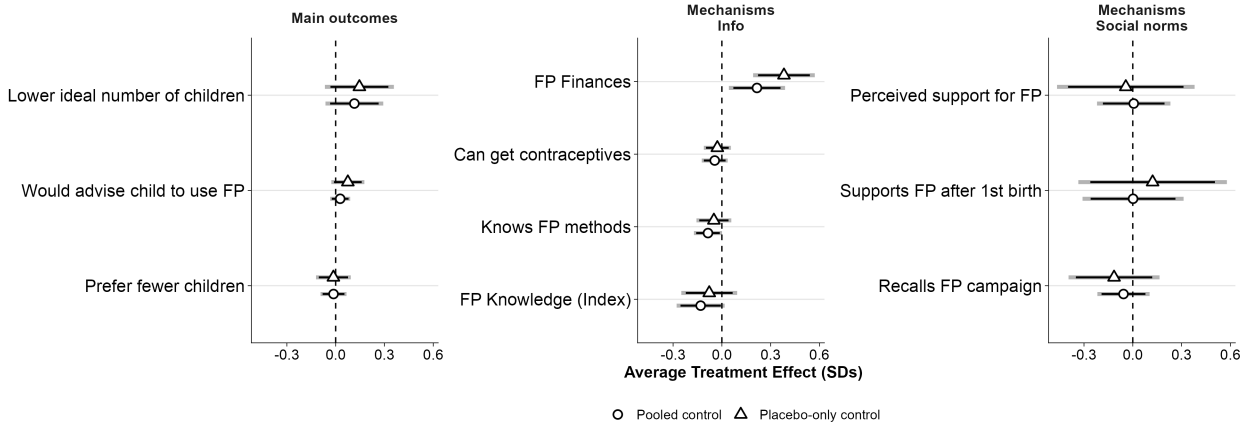



Figure A2: Average Treatment Effects: Family Planning Drama (Robustness)

**Notes:** Circles denote estimates using the pooled control group (placebo plus second treatment arm); triangles denote estimates using the placebo-only control group. Positive values indicate increases relative to the control group, in standardized units. Coefficients based on an OLS regression controlling for gender, age, education, marital status, and the baseline variable where possible. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A9: Average Treatment Effects: FGM Drama (Robustness)

| | Pooled control | | | Placebo-only control | | |
|---|---|---|---|---|---|---|
| Outcome | $\hat{\beta}$ | $p$ | | $\hat{\beta}$ | $p$ | |
| **Oppose FGM** | **0.518** | **0.008** | *** | **0.583** | **0.016** | ** |
| Son can marry uncut woman | 0.125 | 0.227 | | 0.110 | 0.470 | |
| **Would not cut daughter** | **0.176** | **0.005** | *** | **0.187** | **0.024** | ** |
| **FGM Opposition Index** | **0.190** | **0.001** | *** | **0.228** | **0.005** | *** |
| Bride price (16 y/o) | −0.079 | 0.655 | | −0.064 | 0.800 | |

**Bold** = $p \leq 0.10$. Estimates in SDs. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A10: Average Treatment Effects: Family Planning Drama (Robustness)

| | Pooled control | | Placebo-only control | |
|---|---|---|---|---|
| Outcome | $\hat{\beta}$ | $p$ | $\hat{\beta}$ | $p$ |
| FP Knowledge Index | −0.130 | 0.118 | −0.078 | 0.413 |
| Prefer fewer children | −0.013 | 0.753 | −0.014 | 0.808 |
| Lower ideal # of children | 0.115 | 0.237 | 0.146 | 0.232 |
| Would advise child to use FP | 0.028 | 0.400 | 0.075 | 0.205 |

No entries reach $p \leq 0.10$. Estimates in SDs. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

## B.3 Secondary Outcomes: Gender-Based Violence

Figure A3 plots ATE estimates for all GBV-related outcomes for both dramas, using the placebo-only control group. Outcomes span individual attitudes (rejection of wife-beating; beliefs about women reporting violence; opposition to victimization norms), vignette-based behavioral intentions (how respondents say they would react to a cousin disclosing spousal abuse), and community-level intervention norms. Effects are uniformly near zero for both dramas, consistent with the broader pattern of null results on outcomes not directly targeted by the narratives.

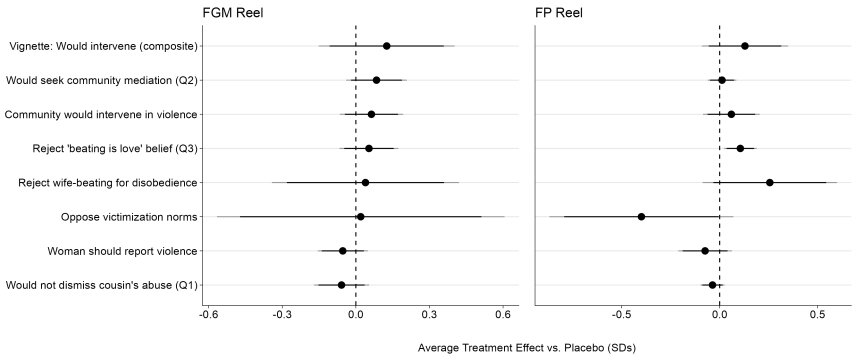



Figure A3: Drama Effects on Gender-Based Violence Outcomes (Placebo-Only Control)
**Notes:** Each panel shows average treatment effects vs. the placebo-only control group, in standardized units. Bold thick lines show 95% CIs; thin lines show 90% CIs. Estimates from OLS with CR2 cluster-robust SEs, controlling for gender, age, education, and marital status (plus a baseline item where available). Vignette items (Q1–Q3) capture how the respondent says they would respond to a cousin disclosing that her husband is beating her.

# C Qualitative Measurement

## C.1 Survey–Interview Alignment

Table A11: Alignment of thematic constructs across survey and interview measures

| Theme | Construct | Survey measure | Interview measure |
|---|---|---|---|
| *FGM* | Opposition / continuation | Whether FGM should stop; willingness to allow a son to marry an uncircumcised woman; whether daughters should remain uncircumcised. | Coded stance toward FGM; support for marriage to an uncircumcised woman; advising against circumcising daughters. |
| | Health harms | Beliefs about health risks, complications, and fertility consequences of FGM. | Mentions of health harms, risks, pain, suffering, or complications from FGM. |
| | Religious justification | Whether FGM is religiously justified or required. | Use or rejection of religious justifications; perceived stance of religious leaders. |
| | Norms / marriageability | Perceived community support for FGM and marriage-related motives. | References to norms, tradition, marriageability, and perceived community support. |
| | Empathy | Reflection on women's suffering or known complications. | Mentions of own or others' negative FGM experiences. |
| *Family Planning* | Support for FP | Preferences over additional children, ideal family size, and advising children to use FP. | Support for FP, smaller families, advising children to use methods, and ideal family size. |
| | Knowledge / methods | Knowledge of methods and where to obtain them. | Mentions of FP methods, knowledge, and service access. |
| | Economic burden | Whether large families strain household resources. | Financial or economic reasons for family planning. |
| | Norms / constraints | Perceived community support for FP; comfort discussing contraception. | Social opposition to FP, norms favoring large families, and family pressure. |
| | Empathy | Reflection on women's burden in large families. | Mentions of personal or others' experiences with FP and family strain. |
| *Early Marriage* | Disapproval / justification | Preferred age of marriage; whether fathers should be discouraged from marrying girls early. | Coded disapproval of early marriage, justification frames, and mentions of pros/cons. |
| | Community norms | Perceived support for girls marrying young. | Whether early marriage is described as common or acceptable locally. |
| *Women's Empowerment* | Education / autonomy | Gender gaps in education and household decision-making. | Support for women's education, autonomy, and resistance to male-only control. |
| | Women's work | Attitudes toward women working and gender equality. | Support for women working inside or outside the home. |
| *Domestic Violence* | Reporting / response | Whether abuse should be reported through formal channels; views on wife-beating. | Support for reporting violence and using formal response channels. |

*Notes:* Survey measures come from structured midline questionnaire items; interview measures come from coded open-ended responses, salience indicators, and post-interview assessments. Rows summarize conceptually aligned constructs rather than individual variables.

Table A12: Baseline Correlations: Qualitative vs. Midline Survey Measures

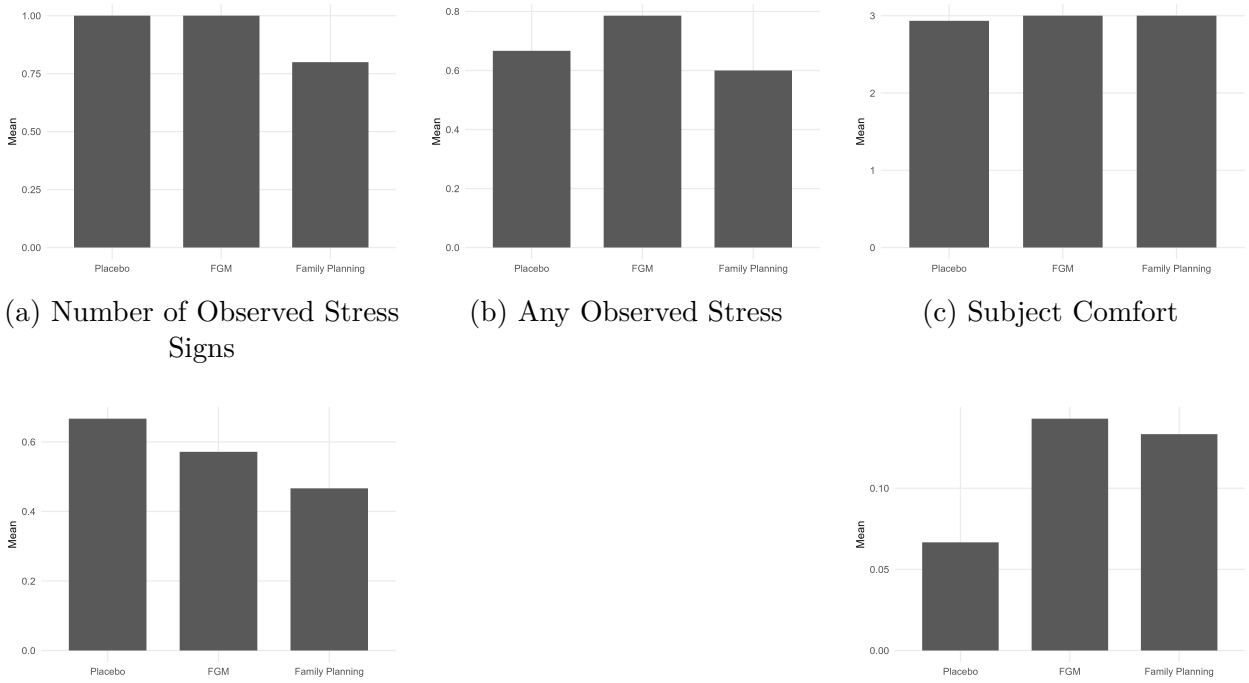
| | $\rho$(Baseline–Qual) | $\rho$(Baseline–Midline) | Qual > Midline |
|---|---|---|---|
| **FGM Attitudes** | | | |
| Would Approve Son Marrying Uncut Woman | 0.261 | 0.858 | ✗ |
| Oppose FGM Continuation | 0.296 | 0.402 | ✗ |
| Presence of Religious Evidence for FGM | 0.187 | 0.196 | ✗ |
| Support Laws Banning FGM | -0.032 | 0.339 | ✗ |
| Aware of FGM Health Harms | 0.499 | 0.645 | ✗ |
| **Family Planning** | | | |
| Salience of Financial Motivations for FP | -0.005 | 0.121 | ✗ |
| Contraceptive Use is Appropriate (Norms) | 0.039 | -0.035 | ✓ |
| Support for Gov. Actions to Limit Family Size | 0.340 | 0.315 | ✓ |
| Believe in Ideal Number of Children (<4) | -0.031 | 0.524 | ✗ |

*Notes:* For rows under *FGM Attitudes*, correlations are computed using only respondents in the placebo and FP-treatment groups. For rows under *Family Planning*, correlations are computed using only respondents in the placebo and FGM-treatment groups. Pearson correlations between baseline responses and (i) qualitative interview-derived measures and (ii) corresponding midline survey outcomes. The final column indicates cases where the baseline–qualitative correlation exceeds the baseline–midline correlation.

## C.2 Interviewer-Coded Outcomes

## C.3 Qualitative Subsample Results

Figures A5–A7 replicate the main survey ATE plots for the subset of respondents ($n = 49$) who were subsequently interviewed face-to-face. The y-axis ordering within each panel matches the full-sample figures exactly.




(a) Number of Observed Stress Signs

(b) Any Observed Stress

(c) Subject Comfort

(d) Most Disengaged Discussing FGM

(e) Most Engaged Discussing FGM

Figure A4: *Behavioral and affective outcomes across discussion conditions.*

*Notes:* Panel (a) shows the number of observed stress signs. Panel (b) shows an indicator for any observed stress. Panel (c) reports the subject's comfort level on a 1–4 scale. Panel (d) shows the percentage of subjects most disengaged while discussing FGM. Panel (e) shows the percentage of subjects appeared most engaged while discussing FGM.

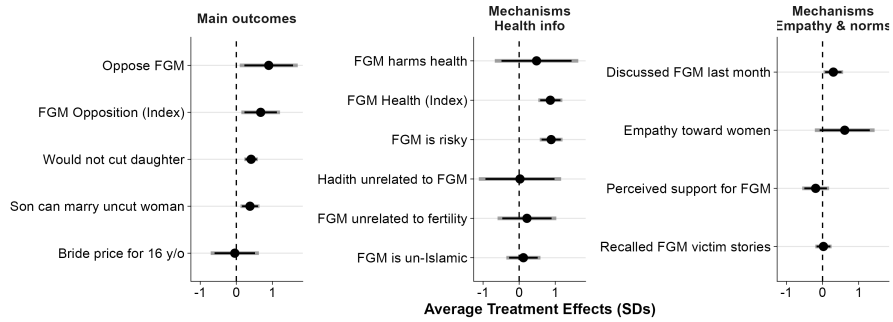



Figure A5: Average Treatment Effects: FGM Drama (Qualitative Subsample)

**Notes:** Pooled control group (placebo + FP arm). Positive values indicate increases relative to the control group, in standardized units. OLS with CR2 cluster-robust SEs. Controls: gender, age, education, marital status, and the baseline item where available. Subset: $n = 49$ respondents who participated in subsequent face-to-face qualitative interviews. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

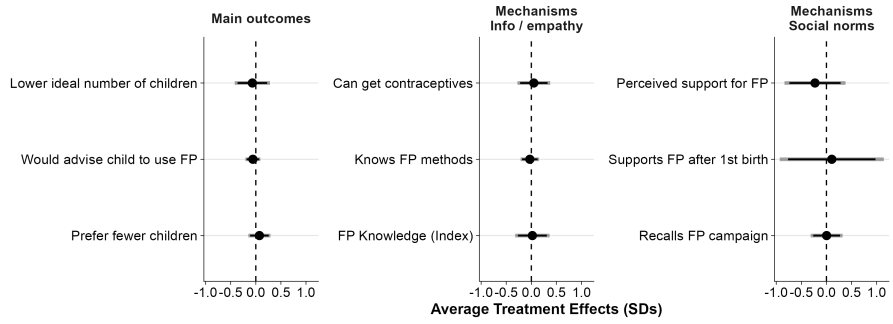



Figure A6: Average Treatment Effects: Family Planning Drama (Qualitative Subsample)

**Notes:** Pooled control group (placebo + FGM arm). Positive values indicate increases relative to the control group, in standardized units. OLS with CR2 cluster-robust SEs. Controls: gender, age, education, marital status, and the baseline item where available. Subset: $n = 49$ respondents who participated in subsequent face-to-face qualitative interviews. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

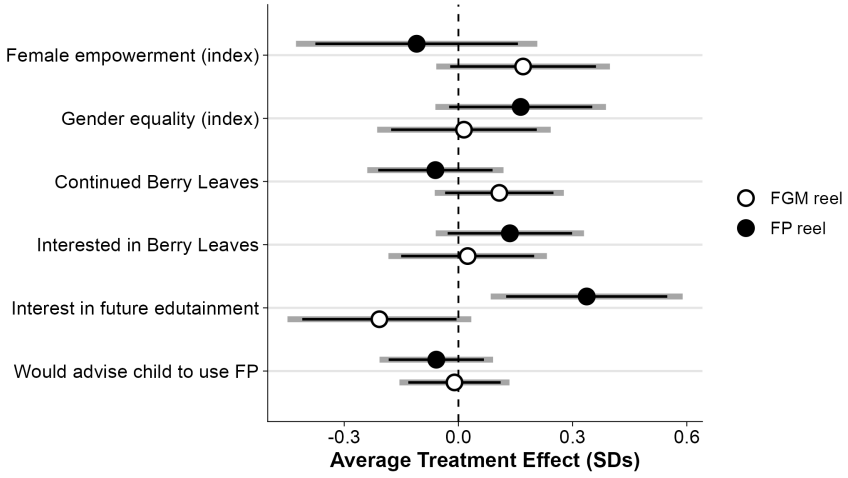



Figure A7: Drama Effects on Broader Gender Attitudes and Media Engagement (Qualitative Subsample)

**Notes:** Open circles denote FGM drama estimates; filled circles denote FP drama estimates. Outcomes and index construction identical to Figure 3. Subset: $n = 49$ respondents who participated in subsequent face-to-face qualitative interviews. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

## C.4 Survey-Aligned Qualitative Outcomes

Table A13: FGM Outcomes

| | (1) FGM | (2) Placebo | (3) Placebo + FP | (4) (1)-(2) | (5) (1)-(3) | (6) N |
|---|---|---|---|---|---|---|
| FGM top-of-the-mind salience | 0.35 | 0.19 | 0.22 | 0.17 | 0.13 | 49 |
| | (0.49) | (0.40) | (0.42) | (0.16) | (0.13) | |
| Anti-FGM stance: Unsupervised assessment (GPT-5) | 8.12 | 7.38 | 6.25 | 0.74 | 1.87** | 49 |
| | (2.15) | (3.22) | (3.42) | (0.95) | (0.91) | |
| Support son marrying uncircumcised girl | 0.94 | 0.81 | 0.80 | 0.13 | 0.14 | 47 |
| | (0.24) | (0.40) | (0.41) | (0.11) | (0.11) | |
| Would advise against circumcising daughters | 0.76 | 0.69 | 0.53 | 0.08 | 0.23 | 49 |
| | (0.44) | (0.48) | (0.51) | (0.16) | (0.15) | |
| Support legal punishment for FGM | 0.71 | 0.44 | 0.41 | 0.27 | 0.30** | 49 |
| | (0.47) | (0.51) | (0.50) | (0.17) | (0.15) | |
| Mention of health harm as FGM concern | 0.44 | 0.69 | 0.68 | -0.25 | -0.24 | 44 |
| | (0.51) | (0.48) | (0.48) | (0.18) | (0.15) | |
| Spontaneuous mention of FGM health risk | 0.65 | 0.75 | 0.62 | -0.10 | 0.02 | 49 |
| | (0.49) | (0.45) | (0.49) | (0.16) | (0.15) | |
| Rejects religious justifications for FGM | 0.65 | 0.47 | 0.37 | 0.18 | 0.28* | 47 |
| | (0.49) | (0.52) | (0.49) | (0.18) | (0.15) | |
| Mentions own negative FGM experience | 0.18 | 0.38 | 0.25 | -0.20 | -0.07 | 49 |
| | (0.39) | (0.50) | (0.44) | (0.16) | (0.13) | |
| Mentions others' negative FGM experiences | 0.29 | 0.50 | 0.31 | -0.21 | -0.02 | 49 |
| | (0.47) | (0.52) | (0.47) | (0.17) | (0.14) | |
| Perceived religious leaders oppose FGM | 0.47 | 0.19 | 0.19 | 0.28* | 0.28** | 49 |
| | (0.51) | (0.40) | (0.40) | (0.16) | (0.13) | |
| Mentions social norms sustaining FGM | 0.82 | 0.81 | 0.88 | 0.01 | -0.05 | 49 |
| | (0.39) | (0.40) | (0.34) | (0.14) | (0.11) | |
| Perceived community support for FGM | 0.53 | 0.69 | 0.60 | -0.16 | -0.07 | 47 |
| | (0.51) | (0.48) | (0.50) | (0.17) | (0.15) | |
| Most engaged discussing FGM | 0.14 | 0.07 | 0.10 | 0.08 | 0.04 | 44 |
| | (0.36) | (0.26) | (0.31) | (0.12) | (0.10) | |
| Most uneasy discussing FGM | 0.57 | 0.67 | 0.57 | -0.10 | 0.00 | 44 |
| | (0.51) | (0.49) | (0.50) | (0.19) | (0.16) | |

Group columns report means with standard deviations in parentheses. Difference columns report differences in means with standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A14: Family Planning Outcomes

| | (1) FP | (2) Placebo | (3) Placebo + FGM | (4) (1)-(2) | (5) (1)-(3) | (6) N |
|---|---|---|---|---|---|---|
| FP top-of-the-mind salience | 0.12 | 0.00 | 0.03 | 0.13 | 0.09 | 49 |
| | (0.34) | (0.00) | (0.17) | (0.09) | (0.07) | |
| Pro-FP Stance: Unsupervised assessment (GPT-5) | 7.56 | 8.56 | 8.27 | -1.00** | -0.71** | 49 |
| | (1.31) | (0.63) | (0.91) | (0.36) | (0.32) | |
| Support for family planning | 0.94 | 0.94 | 0.97 | 0.00 | -0.03 | 49 |
| | (0.25) | (0.25) | (0.17) | (0.09) | (0.06) | |
| Would advise children to use FP methods | 0.94 | 0.94 | 0.97 | 0.00 | -0.03 | 49 |
| | (0.25) | (0.25) | (0.17) | (0.09) | (0.06) | |
| Advise own children on ideal family size | 0.71 | 0.62 | 0.69 | 0.09 | 0.03 | 46 |
| | (0.47) | (0.50) | (0.47) | (0.18) | (0.15) | |
| Support government incentives for smaller families | 0.33 | 0.81 | 0.67 | -0.48*** | -0.33** | 45 |
| | (0.49) | (0.40) | (0.48) | (0.17) | (0.16) | |
| Knows FP differs from birth control | 0.88 | 0.88 | 0.91 | 0.00 | -0.03 | 49 |
| | (0.34) | (0.34) | (0.29) | (0.12) | (0.09) | |
| Mentions health benefits of FP | 0.50 | 0.38 | 0.52 | 0.13 | -0.02 | 49 |
| | (0.52) | (0.50) | (0.51) | (0.18) | (0.16) | |
| Mentions financial reasons for FP | 0.56 | 0.75 | 0.70 | -0.19 | -0.13 | 49 |
| | (0.51) | (0.45) | (0.47) | (0.17) | (0.15) | |
| Mentions experiences with large families and FP | 0.69 | 0.31 | 0.42 | 0.38** | 0.26* | 49 |
| | (0.48) | (0.48) | (0.50) | (0.17) | (0.15) | |
| Perceives community opposition to FP | 0.27 | 0.31 | 0.25 | -0.05 | 0.02 | 47 |
| | (0.46) | (0.48) | (0.44) | (0.17) | (0.14) | |
| Mentions norms favoring large families | 0.88 | 0.94 | 0.94 | -0.06 | -0.06 | 49 |
| | (0.34) | (0.25) | (0.24) | (0.11) | (0.08) | |
| Most engaged discussing FP | 0.33 | 0.53 | 0.38 | -0.20 | -0.05 | 44 |
| | (0.49) | (0.52) | (0.49) | (0.18) | (0.16) | |
| Most uneasy discussing FP | 0.00 | 0.07 | 0.03 | -0.07 | -0.03 | 44 |
| | (0.00) | (0.26) | (0.19) | (0.07) | (0.05) | |

Group columns report means with standard deviations in parentheses. Difference columns report differences in means with standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A15: Female Early Marriage Outcomes

| | (1) FGM | (2) Placebo | (3) Placebo + FP | (4) (1)-(2) | (5) (1)-(3) | (6) N |
|---|---|---|---|---|---|---|
| FEM top-of-the-mind salience | 0.06 (0.24) | 0.06 (0.25) | 0.06 (0.25) | -0.00 (0.09) | -0.00 (0.07) | 49 |
| Anti-FEM Stance: Unsupervised assessment (GPT-5) | 7.82 (2.24) | 9.06 (0.44) | 8.34 (1.60) | -1.24** (0.57) | -0.52 (0.55) | 49 |
| Disapproval of FEM | 0.59 (0.51) | 0.75 (0.45) | 0.55 (0.51) | -0.16 (0.17) | 0.04 (0.15) | 48 |
| Justifies marriage before 18 | 0.41 (0.51) | 0.62 (0.50) | 0.50 (0.51) | -0.21 (0.18) | -0.09 (0.15) | 49 |
| Mentions pros of FEM | 0.76 (0.44) | 0.88 (0.34) | 0.75 (0.44) | -0.11 (0.14) | 0.01 (0.13) | 49 |
| Mentions cons of FEM | 0.94 (0.24) | 1.00 (0.00) | 0.97 (0.18) | -0.06 (0.06) | -0.03 (0.06) | 49 |
| Perceives FEM as common/acceptable in own community | 0.41 (0.51) | 0.38 (0.50) | 0.59 (0.50) | 0.04 (0.18) | -0.18 (0.15) | 49 |

Group columns report means with standard deviations in parentheses. Difference columns report differences in means with standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A16: Women Empowerment Outcomes

| | (1) FGM | (2) FP | (3) Placebo | (4) (1)-(3) | (5) (2)-(3) | (6) N |
|---|---|---|---|---|---|---|
| Support for women's university education | 0.56 (0.51) | 0.86 (0.36) | 0.80 (0.41) | -0.24 (0.17) | 0.06 (0.15) | 45 |
| Pro-Empowerment Stance: Unsupervised assessment (GPT-5) | 6.53 (2.03) | 7.38 (1.67) | 7.81 (1.52) | -1.28** (0.63) | -0.44 (0.56) | 49 |
| Opposes divorce/polygamy for infertility/no sons | 0.35 (0.49) | 0.31 (0.48) | 0.27 (0.46) | 0.09 (0.17) | 0.04 (0.18) | 45 |
| Support for women working to support household | 0.69 (0.48) | 0.80 (0.41) | 0.94 (0.25) | -0.25* (0.14) | -0.14 (0.12) | 47 |
| Support for women working despite financial security | 0.53 (0.51) | 0.53 (0.52) | 0.81 (0.40) | -0.28* (0.16) | -0.28 (0.17) | 48 |
| Acceptability of women working outside home | 1.00 (0.00) | 0.94 (0.25) | 1.06 (0.25) | -0.06 (0.06) | -0.12 (0.09) | 49 |
| Support for reporting domestic violence | 3.00 (1.06) | 2.44 (1.15) | 3.12 (1.20) | -0.13 (0.39) | -0.69 (0.42) | 49 |

Group columns report means with standard deviations in parentheses. Difference columns report differences in means with standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

## C.5 Emotional Reactions Across Groups

Table A17: Emotion prevalence and main composites by arm: FGM

| Emotion | FGM | Placebo | Pooled (Placebo + FP) | FGM - Placebo | FGM - Pooled | N |
|---|---|---|---|---|---|---|
| Compassion/Empathy | 0.294 | 0.500 | 0.375 | -0.206<br>(0.172) | -0.081<br>(0.145) | 49 |
| Hope/Optimism (Change) | 0.353 | 0.312 | 0.156 | 0.040<br>(0.169) | 0.197<br>(0.125) | 49 |
| Pride/Dignity | 0.000 | 0.000 | 0.031 | 0.000<br>(0.000) | -0.031<br>(0.043) | 49 |
| Sadness/Grief | 0.118 | 0.250 | 0.250 | -0.132<br>(0.137) | -0.132<br>(0.122) | 49 |
| Fear/Anxiety | 0.353 | 0.500 | 0.438 | -0.147<br>(0.176) | -0.085<br>(0.150) | 49 |
| Disgust/Aversion | 0.235 | 0.312 | 0.188 | -0.077<br>(0.159) | 0.048<br>(0.123) | 49 |
| Anger/Indignation | 0.471 | 0.500 | 0.438 | -0.029<br>(0.180) | 0.033<br>(0.152) | 49 |
| Resignation/Acceptance (Norm Compliance) | 0.118 | 0.188 | 0.188 | -0.070<br>(0.128) | -0.070<br>(0.113) | 49 |
| Confusion/Uncertainty | 0.059 | 0.125 | 0.062 | -0.066<br>(0.103) | -0.004<br>(0.073) | 49 |
| Neutral/Flat | 0.353 | 0.125 | 0.250 | 0.228<br>(0.148) | 0.103<br>(0.138) | 49 |
| Positive emotions | 0.471 | 0.500 | 0.406 | -0.029<br>(0.180) | 0.064<br>(0.151) | 49 |
| Distressing emotions | 0.529 | 0.750 | 0.625 | -0.221<br>(0.168) | -0.096<br>(0.150) | 49 |

Rows for named emotions report respondent-level means of 0/1 indicators for whether that emotion appears anywhere in the top-5 list. Rows with zero means in all three groups are omitted. Positive emotions and Distressing emotions are respondent-level 0/1 indicators equal to one if any component emotion appears anywhere in the top-5 list. Difference columns report OLS difference-in-means estimates with standard errors shown below in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A18: Emotion prevalence and main composites by arm: FP

| Emotion | FP | Placebo | Pooled (Placebo + FGM) | FP - Placebo | FP - Pooled | N |
|---|---|---|---|---|---|---|
| Compassion/Empathy | 0.375 | 0.250 | 0.364 | 0.125<br>(0.168) | 0.011<br>(0.150) | 49 |
| Hope/Optimism (Change) | 0.062 | 0.250 | 0.242 | -0.187<br>(0.128) | -0.180<br>(0.118) | 49 |
| Pride/Dignity | 0.062 | 0.000 | 0.030 | 0.063<br>(0.062) | 0.032<br>(0.061) | 49 |
| Sadness/Grief | 0.188 | 0.062 | 0.091 | 0.125<br>(0.119) | 0.097<br>(0.101) | 49 |
| Fear/Anxiety | 0.188 | 0.250 | 0.212 | -0.062<br>(0.151) | -0.025<br>(0.125) | 49 |
| Disgust/Aversion | 0.000 | 0.062 | 0.030 | -0.062<br>(0.062) | -0.030<br>(0.044) | 49 |
| Anger/Indignation | 0.312 | 0.375 | 0.303 | -0.063<br>(0.173) | 0.009<br>(0.143) | 49 |
| Resignation/Acceptance (Norm Compliance) | 0.188 | 0.125 | 0.182 | 0.062<br>(0.132) | 0.006<br>(0.120) | 49 |
| Neutral/Flat | 0.500 | 0.375 | 0.364 | 0.125<br>(0.180) | 0.136<br>(0.152) | 49 |
| Positive emotions | 0.500 | 0.312 | 0.485 | 0.187<br>(0.176) | 0.015<br>(0.155) | 49 |
| Distressing emotions | 0.375 | 0.312 | 0.303 | 0.062<br>(0.173) | 0.072<br>(0.145) | 49 |

Rows for named emotions report respondent-level means of 0/1 indicators for whether that emotion appears anywhere in the top-5 list. Rows with zero means in all three groups are omitted. Positive emotions and Distressing emotions are respondent-level 0/1 indicators equal to one if any component emotion appears anywhere in the top-5 list. Difference columns report OLS difference-in-means estimates with standard errors shown below in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A19: FGM Treatment Effects on Relative Emotional Tone Across Topics (FGM vs Family Planning)

| | Treatment effect | | N | |
|---|---|---|---|---|
| Outcome | FGM vs Placebo | FGM vs Pooled | Placebo | Pooled |
| *Panel A. Any Emotion* | | | | |
| FGM negative − FP negative | -0.202<br>(0.258) | -0.046<br>(0.221) | 33 | 49 |
| FGM positive − FP positive | -0.364<br>(0.225) | -0.176<br>(0.208) | 33 | 49 |
| (FGM neg-pos) − (FP neg-pos) | 0.162<br>(0.371) | 0.131<br>(0.305) | 33 | 49 |
| *Panel B. Count* | | | | |
| FGM negative − FP negative | -0.276<br>(0.373) | -0.088<br>(0.304) | 33 | 49 |
| FGM positive − FP positive | -0.430<br>(0.338) | -0.180<br>(0.276) | 33 | 49 |
| (FGM neg-pos) − (FP neg-pos) | 0.154<br>(0.545) | 0.092<br>(0.429) | 33 | 49 |
| *Panel C. Rank-weighted* | | | | |
| FGM negative − FP negative | -0.054<br>(0.093) | -0.031<br>(0.077) | 33 | 49 |
| FGM positive − FP positive | -0.085<br>(0.091) | -0.037<br>(0.075) | 33 | 49 |
| (FGM neg-pos) − (FP neg-pos) | 0.030<br>(0.155) | 0.005<br>(0.122) | 33 | 49 |

Notes: Entries report difference-in-differences-style treatment effects for the FGM arm. For each respondent, outcomes compare emotional content in the FGM topic set relative to the family-planning topic set. In Panel A, the any-emotion specification codes whether at least one emotion from the relevant category appears among the five most salient emotions extracted from the topic-specific transcript. In Panel B, the count specification records the number of distinct emotions in the category. In Panel C, the rank-weighted specification sums salience weights across emotions using weights 5/15, 4/15, 3/15, 2/15, and 1/15 for ranks 1 through 5. Standard errors are shown in parentheses below estimates. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

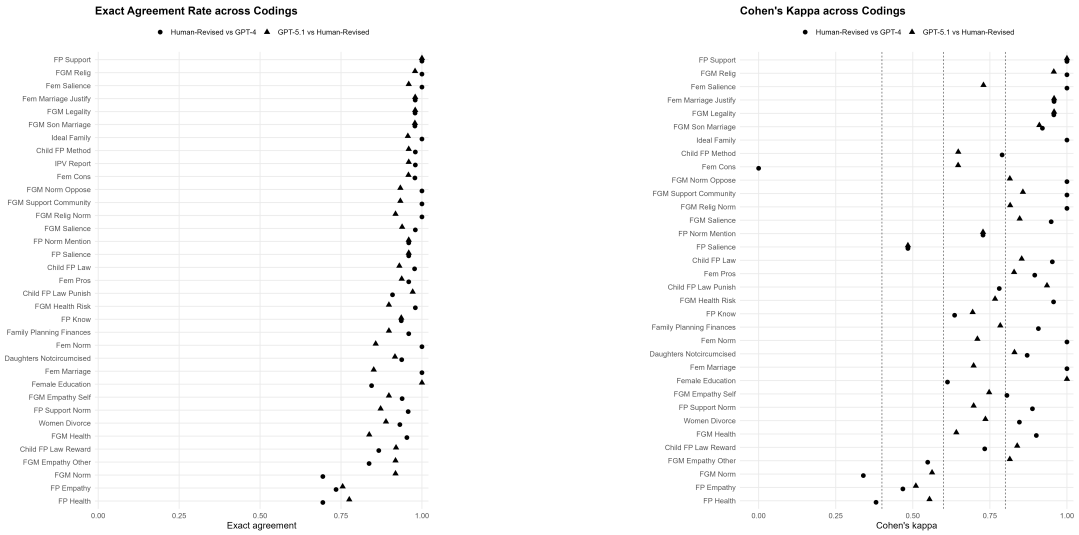
## C.6 LLM Coding Validation

### C.6.1 Survey-Aligned Variables

Table A20: Overall agreement summary across coding versions

| Comparison | Vars | Mean agr. | Med. agr. | Min agr. | Max agr. | Mean $\kappa$ | Med. $\kappa$ | Min $\kappa$ | Max $\kappa$ |
|---|---|---|---|---|---|---|---|---|---|
| Human vs GPT-4 | 35 | 0.94 | 0.98 | 0.69 | 1 | 0.8 | 0.9 | 0 | 1 |
| Human vs GPT-5.1 | 35 | 0.92 | 0.93 | 0.76 | 1 | 0.77 | 0.78 | 0.48 | 1 |

Summary statistics are computed across the 38 substantive roots. Kappa summaries use non-missing kappa values only.




(a) Exact agreement (b) Cohen's kappa

Figure A8: Agreement across coding versions for the 34 substantive qualitative variables.

### C.6.2 Emotional Content

Table A21: Sentiment Analysis: Agreement in Emotion Labels between GPT-4 and GPT-5.1

| Emotion | Prev. GPT-4 | Prev. GPT-5.1 | Agreement | Kappa |
|---|---|---|---|---|
| anger/indignation | 0.194 | 0.316 | 0.816 | 0.526 |
| fear/anxiety | 0.168 | 0.240 | 0.837 | 0.501 |
| neutral/flat | 0.592 | 0.388 | 0.714 | 0.451 |
| sadness/grief | 0.102 | 0.143 | 0.878 | 0.432 |
| hope/optimism (change) | 0.367 | 0.235 | 0.755 | 0.430 |
| compassion/empathy | 0.153 | 0.342 | 0.770 | 0.412 |
| Confusion/uncertainty | 0.005 | 0.020 | 0.985 | 0.395 |
| resignation/acceptance (norm compliance) | 0.204 | 0.133 | 0.827 | 0.386 |
| pride/dignity | 0.046 | 0.061 | 0.934 | 0.347 |
| disgust/aversion | 0.077 | 0.087 | 0.888 | 0.252 |
| Relief/comfort | 0.000 | 0.005 | 0.995 | 0.000 |

Prevalence is the share of respondent-topic observations in which the emotion appears anywhere in the top-5 list. Agreement and kappa compare the two models' binary indicators for whether the emotion appears at all.

Table A22: Alignment between GPT-4 and GPT-5.1 emotion labels by topic

| Topic | N | Exact set | Exact ranked | Top-1 | Jaccard | Weighted Jaccard | Footrule Similarity |
|---|---|---|---|---|---|---|---|
| FGM | 49 | 0.265 | 0.204 | 0.571 | 0.563 | 0.533 | 0.610 |
| Family Planning | 49 | 0.286 | 0.224 | 0.306 | 0.481 | 0.389 | 0.483 |
| Early Marriage | 49 | 0.245 | 0.224 | 0.510 | 0.501 | 0.482 | 0.559 |
| Women's Empowerment | 49 | 0.347 | 0.347 | 0.469 | 0.518 | 0.486 | 0.547 |

Exact set match compares the unordered top-5 emotion sets. Exact ranked match requires the full ranked lists to match exactly. Top-1 is agreement on the top-ranked emotion. Jaccard is set overlap. Weighted Jaccard and Footrule Similarity incorporate rank information.

### C.6.3 Robustness: GPT-4o Results

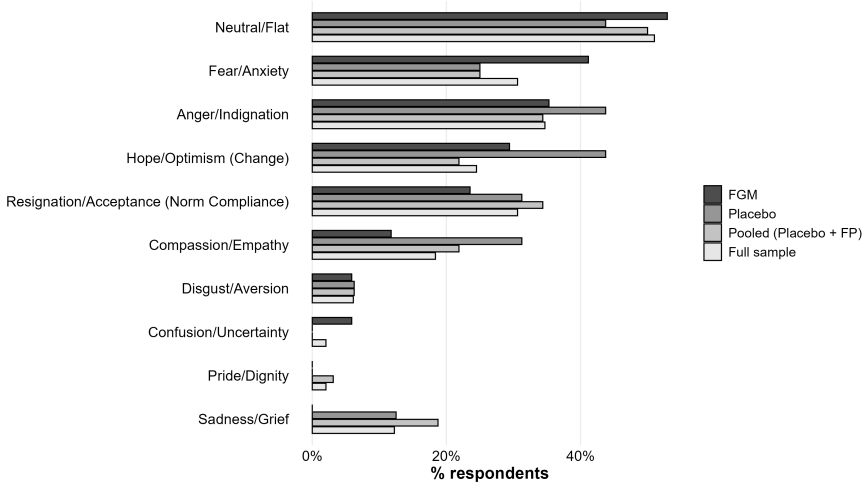



Figure A9: Emotional Content in Interview Transcriptions (GPT-4o): FGM

**Notes:** Figure reports the share of respondents in each treatment arm for whom the listed emotion was included among the five most salient emotions identified by the LLM in the interview transcripts for the FGM question set. Percentages may sum to more than 100% because multiple emotions may be coded for the same respondent. Emotions are ordered by their overall frequency in the treatment condition.

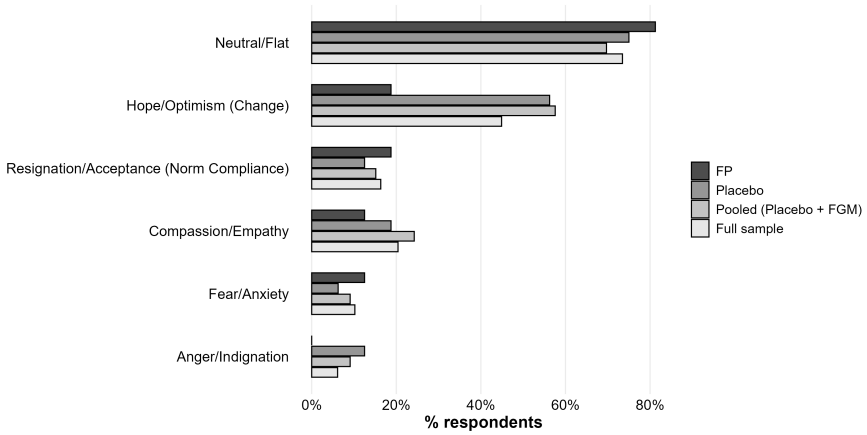



Figure A10: Emotional Content in Interview Transcriptions (GPT-4o): Family Planning

**Notes:** Figure reports the share of respondents in each treatment arm for whom the listed emotion was included among the five most salient emotions identified by the LLM in the interview transcripts for the FGM question set. Percentages may sum to more than 100% because multiple emotions may be coded for the same respondent. Emotions are ordered by their overall frequency in the treatment condition.


(a) FGM

(b) Family Planning

Figure A11: Emotional content in interview transcriptions (GPT-4o)

**Notes:** Figure reports the share of respondents in each treatment arm for whom the listed emotion was included among the five most salient emotions identified by the LLM in the interview transcripts. Panel (a) reports results for the FGM question set, and panel (b) reports results for the family planning question set. Percentages may sum to more than 100% because multiple emotions may be coded for the same respondent. Emotions are ordered by their overall frequency within each question set.

### C.6.4 Robustness: Rank-Weighted Emotions

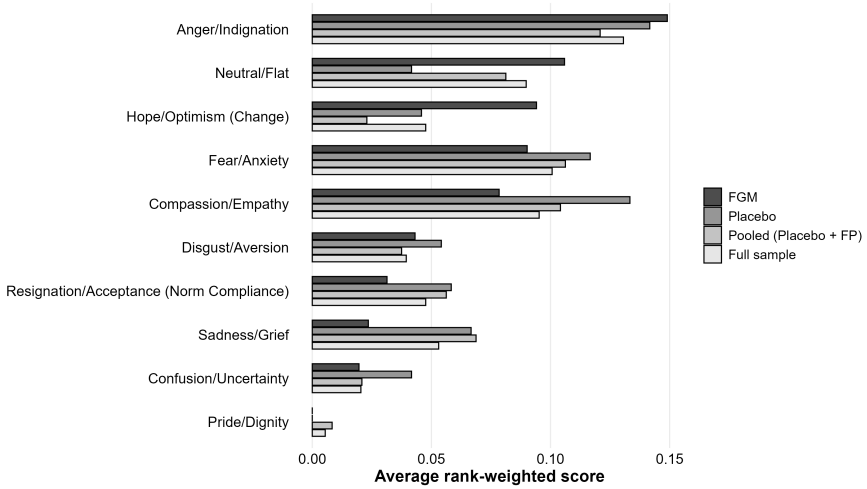



Figure A12: Emotional Content in Interview Transcriptions (Ranked): FGM

**Notes:** Figure reports the share of respondents in each treatment arm for whom the listed emotion was included among the five most salient emotions identified by the LLM in the interview transcripts for the FGM question set. Percentages may sum to more than 100% because multiple emotions may be coded for the same respondent. Emotions are ordered by their overall frequency in the treatment condition.

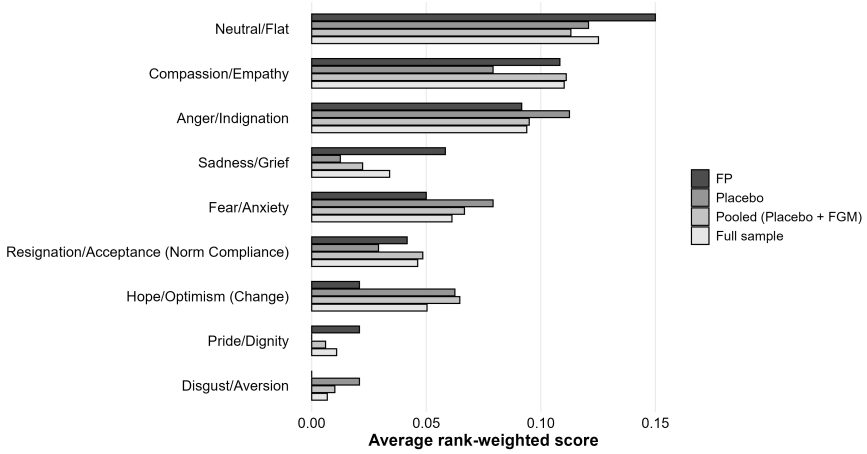
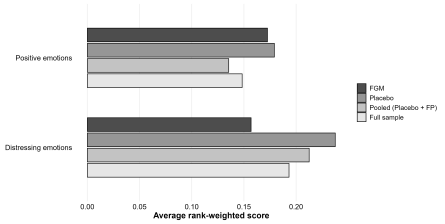
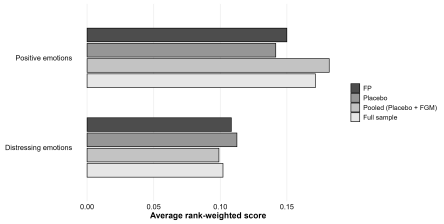



Figure A13: Emotional Content in Interview Transcriptions (Ranked): Family Planning

**Notes:** Figure reports the share of respondents in each treatment arm for whom the listed emotion was included among the five most salient emotions identified by the LLM in the interview transcripts for the FGM question set. Percentages may sum to more than 100% because multiple emotions may be coded for the same respondent. Emotions are ordered by their overall frequency in the treatment condition.




(a) FGM




(b) Family Planning

Figure A14: Emotional content in interview transcriptions (Ranked)

**Notes:** Figure reports the share of respondents in each treatment arm for whom the listed emotion was included among the five most salient emotions identified by the LLM in the interview transcripts. Panel (a) reports results for the FGM question set, and panel (b) reports results for the family planning question set. Percentages may sum to more than 100% because multiple emotions may be coded for the same respondent. Emotions are ordered by their overall frequency within each question set.

# D Sohag Results

Table A23: Effects of FGM Treatment - Sohag Sample

| | FGM Opposition | | FGM Health | | FGM Religion | | Women's Empathy | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treatment effect | 0.271** | 0.171* | 0.373 | 0.184* | 0.314 | 0.260* | -0.050 | -0.169 |
| | (0.060) | (0.078) | (0.182) | (0.080) | (0.213) | (0.134) | (0.304) | (0.142) |
| Comparison group | Placebo | Pooled | Placebo | Pooled | Placebo | Pooled | Placebo | Pooled |
| Observations | 523 | 523 | 522 | 522 | 524 | 524 | 516 | 516 |

Entries are OLS estimates with cluster-robust standard errors in parentheses. All regressions include respondent gender, age, marital status, education, village fixed effects, viewing-date fixed effects, and the prespecified baseline control(s) for each outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A24: Effects of Family Planning Treatment - Sohag Sample

| | FP Support | | FP Knowledge | | FP Finances | | FP Norm | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treatment effect | -0.081 | -0.290 | 0.162 | -0.054 | -0.023 | 0.171 | 0.192 | 0.380 |
| | (0.133) | (0.159) | (0.095) | (0.137) | (0.111) | (0.201) | (0.172) | (0.333) |
| Comparison group | Placebo | Pooled | Placebo | Pooled | Placebo | Pooled | Placebo | Pooled |
| Observations | 525 | 525 | 525 | 525 | 525 | 525 | 523 | 523 |

Entries are OLS estimates with cluster-robust standard errors in parentheses. All regressions include respondent gender, age, marital status, education, village fixed effects, viewing-date fixed effects, and the prespecified baseline control(s) for each outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.