The Long-Term and Spillover Effects of a Radio Drama on HIV Knowledge, Expressed Stigma, Willingness to Disclose, and Political Priorities: Evidence from a Cluster-Randomized Trial in Tanzania

Dylan W. Groves, Alphoncina Kagaigai, Constantine Manda, Beatrice Montano, Savannah Nivin, Bardia Rahmani, Donald P. Green *

August 10, 2026

Abstract

Background. Entertainment-education can influence public health knowledge and attitudes, but evidence on persistence and spillovers is limited.

Methods. We conducted a placebo-controlled, cluster-randomized trial in 30 villages in rural Tanzania. Within 15 matched ward pairs, we randomly assigned villages to a community screening of a radio drama modeling HIV disclosure and help seeking or to a screening of an unrelated drama. We enrolled 1,205 respondents at baseline; 87% attended the screening and 96% completed the endline survey. Participants and enumerators were blinded to assignment. The primary outcomes at 18 months were preregistered indices of HIV knowledge, expressed stigma, willingness to disclose, and HIV policy prioritization. We also surveyed participants' partners, close friends, and teenage children to assess spillovers (secondary outcomes). The study was preregistered on the Open Science Framework.

Results. At 18 months, the HIV drama increased participants' knowledge (d = 0.14), willingness to disclose (d = 0.17), and prioritization of HIV policy (d = 0.17), but not expressed stigma (d = 0.07). These effects were about half as large as those measured two weeks after exposure. Willingness to disclose also increased among the partners (d = 0.16) and teenage children (d = 0.23) of screening attendees. The estimate for partners' policy prioritization was similar (d = 0.16) but less precise. We found no spillover in knowledge or stigma and no effects among friends.

Conclusions. A single group screening produced small changes in HIV-related knowledge and attitudes that remained evident 18 months later. Spillover was selective and limited to household relationships.

*The authors would like to thank the implementing partner, Media for Development International, led by John Ribber. We are grateful to the entire research team at Innovations for Poverty Action, particularly Martin Zuakulu, Rachel Jones, Kyu Khin Gar, Zachary Isdahl, Robert Mwandumbya, Neema Msechu, Gilbert Loshoock, Cosmas Sway, Zinabu Omary, Fatuma Yahaya, Frank Simon, Fadhili Mashaka, Stephano Kizigo, and Saulo Mwakisongo. The studies described in this paper were reviewed and approved by Columbia's Institutional Review Board (protocol AAAR5582) and Tanzania's Commission of Science and Technology (protocol 14528). The trial was preregistered with EGAP on the Open Science Framework (https://osf.io/4tdz2, registered August 10, 2019); the pre-analysis addendum covering the spillover surveys, finalized before spillover data collection began, is available in the study's OSF repository (https://osf.io/c8tfu, posted January 28, 2021) and included in the appendix. These studies were made possible by financial support from Wellspring Philanthropic Fund. Wellspring Philanthropic Fund had no role in the study design, conduct, analysis, or reporting. Corresponding author: Dylan W. Groves (dylanwgroves@gmail.com).

1 Introduction

A growing body of research demonstrates that entertainment education, or "edutainment", can shape knowledge, attitudes, and behaviors related to HIV/AIDS and other public health issues. Specifically, three recent randomized evaluations of entertainment-education dramas focused on HIV demonstrate improvement in HIV knowledge and treatment (Banerjee et al. 2019), a decrease in sexual risk-taking behavior (Basaran et al. 2019), and an increase in HIV disclosure intentions and policy prioritization [REDACTED for review].

But two questions about the influence of entertainment-education interventions to combat HIV/AIDS have received comparatively less attention. First, do the effects of narrative media persist beyond several days? Second, how widely do the effects of the drama diffuse beyond the original audience into the audience's familial and social networks? A recent meta-analysis of 77 experimental evaluations of narrative media interventions found limited reporting of long-term effects: the majority recorded participant responses immediately, while only five studies followed participants for 12 months or longer (Banerjee et al. 2015; Carey et al. 2015; Hu et al. 2018; Bjorvatn et al. 2020; Semakula et al. 2020), and just two of those extended follow-up to 24 months (Hu et al. 2018; Bjorvatn et al. 2020). None of the five studies with follow-up of 12 months or longer investigated HIV/AIDS-related issues, and none measured spillover effects (Rahmani et al. 2025).

The persistence and diffusion of the effects of narrative media interventions have important implications for their scalability and cost-effectiveness. If treatment effects fade quickly, then entertainment-education campaigns may need ongoing reinforcement. On the other hand, if effects are durable and diffuse through social networks, then the cost-effectiveness of entertainment-education interventions may be far greater than previously estimated. Advocates of entertainment-education consistently point to neighborhood spillovers as an important force multiplier for their interventions (Mohammed 2001; Sharan and Valente 2002).

This study addresses both gaps by tracking the effects of an influential HIV-focused radio drama over an 18-month period and examining spillover to three distinct social groups: the original listeners' partners, friends, and children. The radio drama was designed to increase awareness of HIV treatments, reduce stigma, and encourage status disclosure. The original study randomly assigned 30 villages to hear either the HIV-focused radio drama or a placebo drama about an unrelated topic, and surveyed 1,205 respondents

before treatment (baseline) and two weeks post-treatment (midline). The present study builds on this RCT by surveying the original respondents 18 months post-treatment (endline) and by extending data collection to their partners, close friends, and teenage children.

We examine the drama's effects on four families of pre-registered outcomes. *HIV knowledge* captures factual knowledge about HIV transmission and antiretroviral (ARV) treatment. *Expressed stigma* captures stated acceptance of persons living with HIV in social and economic life. *Willingness to disclose* captures respondents' stated willingness to share a hypothetical HIV-positive diagnosis with family members, friends, and coworkers. *HIV policy prioritization* captures the priority respondents assign to HIV/AIDS treatment among competing policy goals. We measure each outcome among primary respondents 18 months after the intervention, and we assess diffusion by measuring corresponding outcomes among primary respondents' partners, close friends, and teenage children.

2 Literature Review & Theoretical Background

2.1 Entertainment-Education

Embedding messages within compelling narratives has become a popular strategy for promoting public health and social change (Paluck and Green 2009; Frank and Falzone 2021). At a practical level, the ability of narrative dramas to attract and retain the attention of audiences makes them a useful vehicle for delivering public service messages (Green et al. 2024). Narrative entertainment advocates propose two further reasons why we might expect narrative messages to be especially persuasive. First, audiences may learn vicariously by observing and identifying with characters who model certain behaviors (Bandura 2004). Second, audiences may evince less skepticism toward a persuasive message when they are absorbed in an engaging narrative (Green and Brock 2000), creating openings for attitude change that remain closed to more didactic approaches (Slater and Rouner 2002; Jones and Paris 2018).

Experimental evidence is consistent with these theoretical propositions. A recent comprehensive meta-analysis of experimental studies of narrative messaging found that narrative interventions consistently generate effects on knowledge, attitudes, and behavioral intentions (Rahmani et al. 2025). A meta-analysis focused specifically on entertainment-education and health outcomes drew the same conclusion (Doré et al. 2025).

Entertainment-education has been an especially popular public health tool in lower and middle-income countries (LMICs), where field experiments have evaluated dramas promoting family planning in Tanzania

(Vaughan and Rogers 2000; Rogers et al. 1999), consumption of iron-fortified salt in India (Banerjee et al. 2015), and disclosure and reporting of intimate partner violence in Uganda (Green et al. 2020). Effect sizes in such naturalistic studies are typically smaller than in laboratory settings, perhaps because lab studies tend to measure outcomes immediately after exposure (Rahmani et al. 2025).

2.2 Persistence and Diffusion of Narrative Effects

Two theoretical questions motivate our research design: whether effects persist after a single exposure and whether they diffuse beyond direct audiences.

Theories of persuasion and memory provide reasons to expect effects to decay. For example, the elaboration likelihood model predicts that attitudes are more durable when they are formed through extensive processing (Cacioppo et al. 1986). Because audiences of narrative media typically focus on characters and plot rather than the underlying message, changes in their attitudes and beliefs may be less persistent.

However, some theories of narrative persuasion suggest that effects may endure. Social cognitive theory holds that audiences store behaviors modeled by narrative dramas as scripts in their long-term memory and retrieve those scripts when they encounter similar situations (Bandura 2004). Transportation theory and its extensions propose that absorption links a story's message to memorable characters and events (Green and Brock 2000; Slater and Rouner 2002). Some experimental studies find that narrative effects persist or even strengthen after a delay (Appel and Richter 2007; Dahlstrom 2012).

The second question concerns diffusion. The two-step flow and related theories of social diffusion propose that media influence spreads when audience members transmit and interpret messages through interpersonal communication (Katz 1957; Southwell and Yzer 2007, 2009). Diffusion may depend on the nature of the content: simple and uncontroversial messages may travel more widely than complex or sensitive ones (Centola and Macy 2007). Diffusion may also depend on the relationship: trust and frequent contact may be required for the spread of sensitive information and attitudes (Boulay et al. 2002; Murphy 2004).

However, the few existing experimental studies generally find little transmission to the family members or friends of entertainment-education audiences (Wilke et al. 2020). The closest precedent is Banerjee et al. (2019)'s evaluation of the Nigerian television drama *MTV Shuga*, which produced marginal spillover of HIV knowledge, but not attitudes or behavior, to viewers' friends. The present study extends this literature by measuring the effects of a radio drama at both two weeks and 18 months and by examining spillover separately among partners, close friends, and teenage children.

2.3 The Present Study: Aims and Hypotheses

The present study pursues two pre-registered aims. The first aim is to estimate the effects of an HIV-focused radio drama on primary respondents' HIV knowledge, expressed stigma, willingness to disclose, and policy prioritization 18 months after the intervention, thereby assessing the persistence of effects documented at the two-week midline [REDACTED for review]. The second aim is to estimate whether these effects diffuse through respondents' close social networks, measured through separate surveys of respondents' partners, close friends, and teenage children.

We test five pre-registered directional hypotheses: we expect the drama to increase HIV knowledge (H1), reduce expressed stigma (H2), increase willingness to disclose an HIV-positive status (H3), and increase HIV policy prioritization (H4) among primary respondents at the 18-month endline, and we expect corresponding effects among partners, close friends, and teenage children of primary respondents (H5). The four indices measured on primary respondents constitute the study's primary outcomes; spillover outcomes are secondary.

3 Research Design & Methods

3.1 Study setting

The study was conducted in the Tanga region of northeastern Tanzania. Three features of the study setting are especially relevant to the questions that motivate this paper.

First, HIV is prevalent and disclosure is stigmatized. The 2016–2017 Tanzania HIV Impact Survey found that 8.7% of rural households nationwide include at least one HIV-positive member, with Tanga's prevalence comparable to the national rural rate (National Ministry of Health, Community Development, Gender, Elderly, and Children 2017). Nearly half (43%) of HIV-positive individuals in the region were unaware of their status. National ARV coverage among adults living with HIV reached 75% by 2019 (Avert 2020), but coverage in rural communities is believed to be much lower, in part because rural Tanzanians fear that seeking treatment will expose their status (Audi et al. 2021; Sanga et al. 2019). The drama addresses these barriers directly: its protagonist learns to accept his HIV-positive status, discloses to his family and coworkers, and seeks treatment.

Second, household communication about HIV occurs but is far from universal. A study in coastal Tanzania found that roughly two in five adults reported discussing HIV with their most recent sexual partner (Fonner et al. 2020), and ethnographic research in rural Tanzania finds that most parents discuss

sexual and reproductive health with their children (Wamoyi et al. 2010). Prior evaluations of Tanzanian radio dramas report that audiences discussed program themes with spouses and family members at home (Vaughan and Rogers 2000; Rogers et al. 1999).

Third, social ties beyond the household are weaker than stereotypes of rural Sub-Saharan Africa suggest. Rural Tanzanian villages (*kijiji*) typically comprise several hundred to over a thousand residents divided into hamlets (*kitongoji*), and few villagers in rural Tanga report knowing all or even most of their fellow villagers (Mdee and Mushi 2021; Montano 2026). Moreover, qualitative studies of rural Tanzania report widespread reluctance by villagers to discuss HIV status beyond close family members (Audi et al. 2021; Sanga et al. 2019).

3.2 Intervention

The treatment is a 110-minute abridged version of the serialized Swahili radio drama *Wahapahapa* (*Those Who Belong Here*), written and produced by Media for Development International. The abridged drama follows a young musician named Ray as he learns to accept his HIV-positive status, disclose his status to his family and friends, and obtain life-saving treatment. A detailed description of the reduced-length version of the drama appears in subsection E.2.

Placebo villages were presented with a 110-minute abridged version of *Tamapendo*, a separate serialized Swahili radio drama produced by UZIKWASA, which follows a young girl as she tries to escape an early forced marriage arranged by her father. The placebo does not discuss any issues related to HIV.

The dramas were presented in each village at community listening events. Each screening was overseen by a three-person field team of trained field officers who monitored attendance, verified that the correct drama was played in full, and distributed refreshments. The research team members did not discuss the content of the dramas before, during, or after the listening sessions. The implementation teams documented participant reactions and discussions in every village, which the research team reviewed for signs of participant distress and other unintended harms.

3.3 Sample Selection & Randomization

Site Selection. We conducted the study in 30 rural villages distributed evenly across 15 wards. We employed a placebo-controlled, parallel-group design with cluster randomization at the village level, allocating villages 1:1 to *Wahapahapa* or *Tamapendo* within matched ward-pairs. Villages were eligible if they were sufficiently rural and remote (at least 70km from Pangani town, at least 8km from any town

or city, and no main or secondary road running through the village polygon), of adequate size (100–1,000 households within the sampling radius), and sufficiently distant from other sample villages (at least 4km apart) to limit treatment spillover across villages.

Respondent Sampling. In each village, we randomly selected 40 households with at least one household member aged 18 to 65 years old for participation. To be eligible, households needed to reside within a researcher-selected radius of the village center for at least six months. The village radius was selected to encompass the residential settlement area of the village (see pre-registration in subsection E.3). Households were drawn from a full census within the village radius with no minimum-distance restriction between selected households. Because the dramas were delivered at village-level community listening events rather than through home radio broadcasts, radio ownership was not an eligibility criterion. Within each selected household, one eligible member was randomly selected for participation. Respondents were evenly split between male and female.

As part of our 18-month endline, enumerators also interviewed the partners of all respondents (N = 816). Partners were included in the sample if they were married or co-habitating with the original respondent at baseline. Two to three months after the 18-month endline, we interviewed all of the original respondents' children between the ages of 13 and 18 (N = 554) and one of the closest friends mentioned by the respondent during the endline survey. To construct the friends sample, we asked each respondent at endline to list their three closest friends in the village; after excluding friends who had already been included in the original sample, we randomly selected one of the remaining friends for inclusion in the friends survey (N = 1,109).

Random Assignment. We grouped villages into 15 same-ward pairs and randomly assigned one village in each pair to the treatment and one to the placebo condition. Random assignment was performed using Stata by a research team member not involved in enrollment or surveying. The randomization code is included in the replication materials. Implementation teams traveled to each village carrying the randomly assigned audio recording in a sealed envelope which was not opened until the listening event began.

Sample Size. The sample of 30 villages (15 matched pairs) and 40 respondents per village was chosen to balance budget constraints and a minimum detectable effect target of roughly 0.2 standard deviations on primary outcomes. Because the intracluster correlation was unknown prior to the study, no definitive ex ante power calculation was possible. Appendix I reports minimum detectable effects for the primary and spillover outcomes given the realized design.

3.4 Data Collection

The surveys were rolled out consecutively across wards, so that the treatment and placebo pair in each ward received the baseline survey, the audio screening, and the endline and spillover surveys at similar points in time. Baseline interviews were conducted between July 24 and August 29, 2019, and the village screenings were held between July 25 and September 10, 2019. The midline survey followed 13 to 16 days after each village's screening (August 7 to September 12, 2019), and the endline survey took place approximately 18 months after the screenings (December 7, 2020 to March 17, 2021); respondents were surveyed at both waves regardless of screening attendance. Village-level treatment assignment was not disclosed to enumerators at any time.

Surveys lasted approximately 45-60 minutes and were conducted in Swahili by trained Tanzanian enumerators employed by Innovations for Poverty Action. Enumerators were college-educated, IRB-certified native Swahili speakers with prior household-survey experience and completed a week of training and field-based piloting that included protocols for administering sensitive questions about HIV status and stigma. Enumerators were gender-matched to respondents for all surveys.

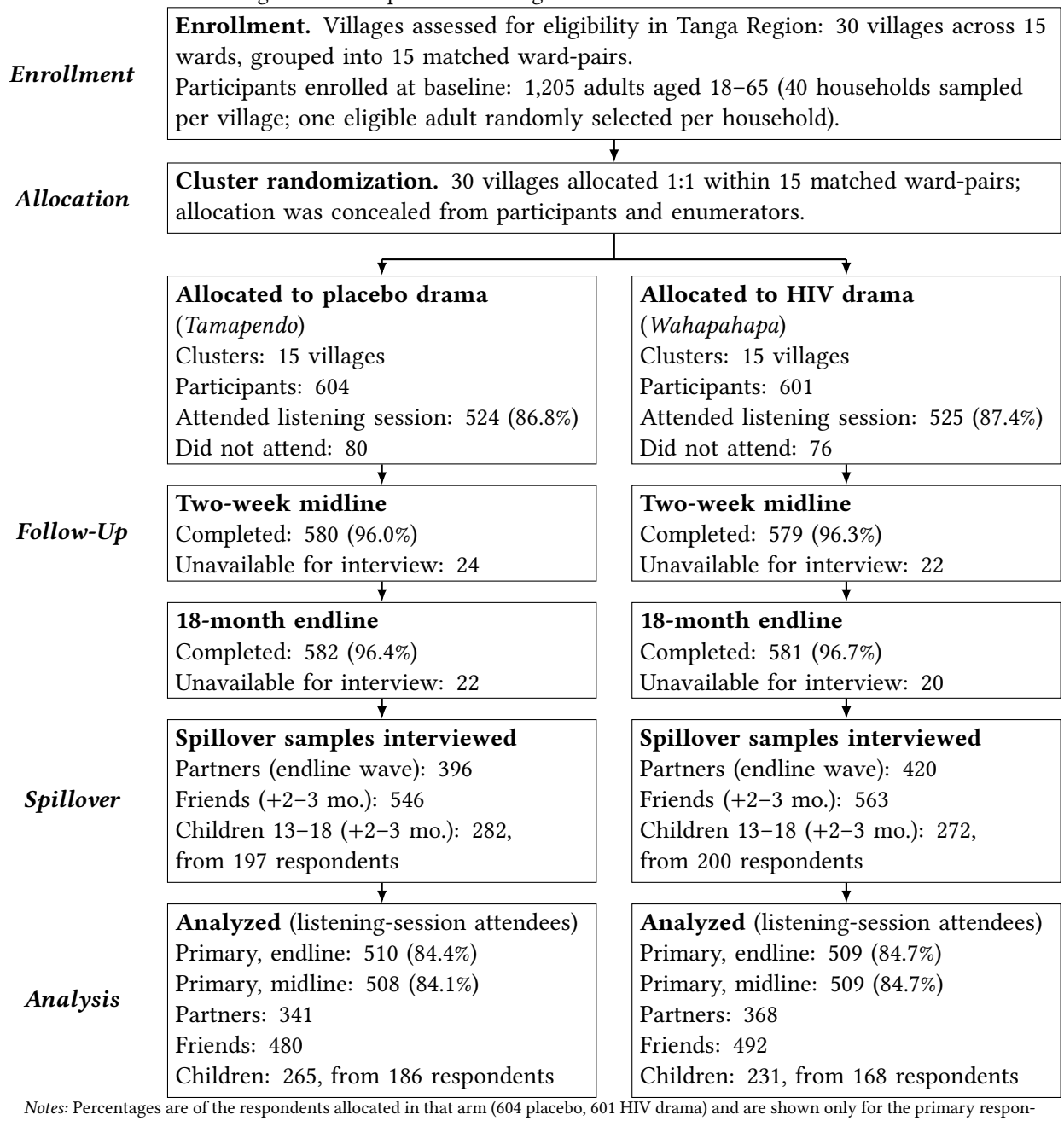
The surveys of friends and teenage children were conducted two to three months after the 18-month endline (January 27 to March 19, 2021). We conducted the partner surveys as part of the same village visit as the endline survey but delayed the friend and child interviews so that we could construct the roster of children and friends from the endline survey. One implication of this approach is that spillover effects on friends and children are measured 20–21 months post-treatment. Data collection concluded after the planned waves; the trial was not stopped early.

We collaborated with Tanzanian research partners to craft culturally appropriate question wording, ensuring questions about HIV status and stigma avoided implicit pressure for status disclosure and minimized potential for emotional distress among respondents. Villagers were not otherwise involved in the design, conduct, or reporting of this study.

We did not collect respondents' HIV serostatus, through either biomarker testing or self-report: the intervention targeted disclosure attitudes, stigma, and policy priorities in the general population, and soliciting serostatus in a study of stigma would have risked the very harms the drama was designed to reduce.

Figure 1 summarizes the flow of clusters and participants through each stage of the trial.

Figure 1: Participant flow through the cluster-randomized trial



Notes: Percentages are of the respondents allocated in that arm (604 placebo, 601 HIV drama) and are shown only for the primary respondents, whose counts are subsets of that denominator. The spillover samples are reported as counts: partners, friends and teenage children exist for only a subset of respondents, so a percentage of allocated respondents would conflate eligibility with response. No clusters were lost after allocation and all enrolled participants remained in their assigned cluster; listening-session attendance is the compliance margin, not a loss of allocated units. Midline and endline completion is reported among all baseline respondents irrespective of attendance, so the endline count exceeds the midline count where respondents missed the two-week wave but were re-contacted at 18 months. Partners (N = 816) were interviewed in the endline wave; friends (N = 1,109) and teenage children (N = 554) were rostered at endline and interviewed two to three months later, i.e. 20–21 months post-treatment. Per the pre-analysis plan, primary analyses are restricted to listening-session attendees; full randomized-sample estimates appear in the Online Appendix.

3.5 Outcome Measures

In line with pre-registration, we measured outcomes related to HIV knowledge, stigma, disclosure, and policy prioritization. The full question text appears in Table A20; surveys of partners, friends, and children administered subsets of the questions administered to the primary respondents to reduce survey length and ensure age appropriateness.

Where multiple items relate to a common outcome family, we pre-registered aggregate indices (equally weighted means of each respondent's available items) to summarize treatment effects and limit multiple comparisons (Kling et al. 2007). Table A21 reports index composition and internal consistency by sample.

HIV Knowledge. We measured respondents' knowledge about the spread and treatment of HIV, centering on respondents' knowledge of ARV treatments and HIV transmission mechanisms.

Willingness to Disclose. We assessed respondents' support for HIV disclosure with questions about respondents' willingness to share their HIV status with a friend and coworker, and another question asking whether respondents would want a family member to disclose their HIV status.

Expressed and Perceived HIV Stigma. We assessed respondents' expressed and perceived HIV stigma by asking them whether they would be willing to sit next to someone with HIV/AIDS and if a worker with HIV should be fired. Enumerators also asked respondents whether they perceived that most others in their community believed that an individual should be fired if they are HIV-positive.

HIV Policy Prioritization. We assessed respondents' prioritization of HIV policy by asking them to vote in a hypothetical local election in which one candidate promised to "increase the availability of medicine for HIV/AIDS" while the other candidate ran on a platform to either reduce crime or improve the quality of roads.

We also asked respondents how they would rank responding to HIV/AIDS among other goals for their village. Enumerators presented a set of nine cards, each displaying the name and an image illustrating a village goal (including increasing the availability of medicine for HIV/AIDS), which respondents ranked from most to least important to their community. We coded the outcome as 1 if the respondent ranked HIV/AIDS response as their top priority.

3.6 Statistical Analyses

The study design, directional hypotheses, outcome measures, and analytical strategy were preregistered on the Open Science Framework prior to endline data collection. The complete pre-registration and

addendum appear in subsection E.3. The registry identifier is withheld for anonymous review.

The main text reports linear multilevel mixed models that account for the nesting of respondents within villages, estimated separately for each respondent sample (primary respondents, partners, friends, and teenage children). For respondent i in village j, we denote the measured outcome as Y_{ij} and village treatment assignment as T_j (coded 1 for the treatment HIV drama, 0 for the placebo drama). We specify the following two-level model:

$$Y_{ij} = \beta T_j + \gamma_1 ward_{1j} + \gamma_2 ward_{2j} \ldots + \gamma_k ward_{kj} + u_j + e_{ij}$$

This specification models outcomes as a linear combination of treatment status, a full set of ward fixed effects (corresponding to the k randomization blocks, i.e., the matched ward pairs), a village random intercept $u_j \sim N(0, \sigma_u^2)$, capturing the interdependence of respondents within the same village, and an independent individual error term $e_{ij} \sim N(0, \sigma_e^2)$. σ_u^2 and σ_e^2 are the village-level and residual variance components reported in the results tables. In the teenage-children sample we add a household random intercept nested within village. Models are estimated by restricted maximum likelihood, and standard errors and degrees of freedom for the treatment coefficient use the Kenward-Roger small-sample correction. In line with our pre-registration, we calculate exact p-values through randomization inference under the sharp null hypothesis of zero treatment effect for every unit.

Our primary parameter of interest, β, is the average treatment effect among participants who attended a screening. It corresponds to the complier average causal effect (CACE) under the assumption that attendance did not depend on treatment assignment. In subsection D.4, we also estimate treatment effects for the whole sample. The results are smaller in magnitude but nearly identical in pattern to the screening-attendee results (subsection 4.2).

The pre-registered design-based specification appears in Appendix L, and a side-by-side comparison in Appendix M shows that no substantive conclusion differs across specifications. No covariates beyond the randomization-block fixed effects enter the primary specification (covariate-adjusted robustness checks appear in Appendix J).

4 Results

4.1 Preliminary Results

4.1.1 Balance

Table 1 reports basic demographic characteristics of the analysis samples. Endline respondents, partners, and friends are similar: average age for the three groups was approximately 38 years, the samples were evenly split between male and female, and about two-thirds of all samples were Muslim. Children of respondents were 15 years old on average.

As expected, random assignment is not significantly correlated with pretreatment characteristics of screening attendees, endline respondents, or endline respondents' partners, friends, or children. Analysis of pretreatment balance between treatment and placebo groups across all baseline variables appears in Appendix Table A4. Of 50 pretreatment covariates measured at baseline, only one covariate (2.0%) shows a difference between treatment and placebo attendees that is statistically significant at the 0.05 level, and only two covariates (4.0%) are statistically significant at the 0.1 level.

Table 1: **Pretreatment Characteristics and Covariate Balance, by Respondent Type**

	Placebo	Treatment	RI p-value
Panel A: Endline Respondents			
Age	38.23 (12.00)	38.71 (11.98)	0.648
Female	0.52 (0.50)	0.53 (0.50)	0.794
Reached Standard 7	0.76 (0.43)	0.76 (0.42)	0.891
Muslim	0.64 (0.48)	0.65 (0.48)	0.863
Times/Week Visit Church/Mosque	5.06 (10.62)	4.72 (10.15)	0.759
N	524	525	
Panel B: Partners			
Age	37.86 (11.32)	39.53 (11.64)	0.072*
Female	0.52 (0.50)	0.51 (0.50)	0.931
Muslim	0.71 (0.45)	0.65 (0.48)	0.346
Times/Week Visit Church/Mosque	6.38 (11.72)	6.14 (11.53)	0.786
N	341	368	
Panel C: Friends			
Age	37.28 (12.02)	37.27 (12.29)	0.984
Female	0.52 (0.50)	0.54 (0.50)	0.224
Reached Standard 7	0.82 (0.39)	0.78 (0.41)	0.301
Muslim	0.67 (0.47)	0.64 (0.48)	0.671
Times/Week Visit Church/Mosque	4.79 (10.15)	4.27 (9.59)	0.602
N	480	492	
Panel D: Children			
Age	15.19 (1.74)	15.05 (1.64)	0.470
Female	0.46 (0.50)	0.44 (0.50)	0.749
Reached Standard 7	0.76 (0.43)	0.72 (0.45)	0.633
Muslim	0.74 (0.44)	0.67 (0.47)	0.150
Times/Week Visit Church/Mosque	6.42 (10.87)	5.33 (10.28)	0.281
N	264	231	

Note: * p < 0.10, ** p < 0.05, *** p < .01. Each panel reports demographic characteristics of one analysis sample, separately by treatment arm; cells contain means with standard deviations in parentheses. The final column reports two-sided randomization-inference p-values testing the null hypothesis of no difference between treatment and placebo villages, computed from 2,000 permutations of village-level treatment assignment within matched ward pairs. Characteristics of endline respondents are measured in the baseline survey; characteristics of partners, friends, and teenage children are measured in their respective surveys, and each panel includes only interviewed members of that sample. Balance tests across all 50 baseline covariates appear in Table A4. Full question text appears in the pre-registration (subsection E.3).

4.1.2 Attrition and Compliance

The assigned drama was played in full at every village, as confirmed by the field officers overseeing each screening. Compliance rates were high and were balanced across the treatment and placebo groups (see Table A22). The initial sample comprised 1,205 respondents who completed a baseline survey and were invited to attend a village screening. The attendance rate was 86.8% in the placebo group and 87.4% in the treatment group (two-tailed RI p-value = 0.86).

Among baseline respondents, 96% completed the midline survey and 96% completed the endline sur-

vey in both the treatment and placebo groups (two-tailed RI p-values = 0.81 and 0.896). In the placebo group, 66% of respondents had a partner who participated in the endline survey; in the treatment group 70% of respondents had a partner who did so (two-tailed RI p-value = 0.268). Enumerators successfully identified and surveyed a close friend of 90% of placebo group respondents and of 93% of treatment group respondents (two-tailed RI p-value = 0.353), and surveyed an average of 0.69 teenage children per placebo group respondent and 0.70 per treatment group respondent (two-tailed RI p-value = 0.814).

We also compared endline respondents (N = 1,163) with dropouts (N = 42) across sociodemographic characteristics and baseline HIV-related variables (Table 2). Dropouts resembled completers on baseline HIV awareness and attitudes and on most demographic characteristics, although they were less likely to have reached Standard 7 (57% versus 77%, $p = 0.005$).

Table 2: **Endline Completers versus Dropouts**

	Completers	Dropouts	Difference	p-value
Age	37.93 (12.01)	37.52 (11.82)	-0.40	0.832
Female	0.50 (0.50)	0.43 (0.50)	-0.08	0.235
Reached Standard 7	0.77 (0.42)	0.57 (0.50)	-0.20	0.005
Muslim	0.65 (0.48)	0.74 (0.45)	0.08	0.183
Times/Week Visit Church/Mosque	4.89 (10.37)	3.68 (8.97)	-1.21	0.386
Heard of HIV/AIDS (baseline)	1.00 (0.04)	0.98 (0.15)	-0.02	0.356
Comfortable on Bus with HIV+ Person (baseline)	0.91 (0.29)	0.88 (0.33)	-0.02	0.734
HIV+ Safe to Work with Children (baseline)	0.45 (0.50)	0.40 (0.50)	-0.04	0.598
N	1163	42		

Note: Cells report means with standard deviations in parentheses among the 1,205 baseline respondents, by endline completion status. p-values are from regressions of each characteristic on an indicator for dropping out, with standard errors clustered by village. This comparison is descriptive: it characterizes who was lost to follow-up rather than testing an experimental design property. Baseline HIV items are coded so that higher values indicate greater awareness or acceptance.

4.2 Main Results

To convey the substantive magnitude of treatment effects, we report standardized effect sizes (the treatment effect divided by the placebo group standard deviation, with small-sample correction, as reported in each results table; standardized effects for all outcomes appear in Appendix K). For the four primary indices, standardized effects at the 18-month endline are 0.14 (knowledge), 0.07 (stigma), 0.17 (disclosure), and 0.17 (prioritization).

Exploratory analyses find very little evidence of treatment-effect heterogeneity on standard demographic variables, including across gender, religion, education, and housing quality (Appendix H).

4.2.1 Knowledge and Beliefs

Eighteen months after the treatment, respondents in the treatment group were 3 percentage points more likely to know that ARVs or other medicines help persons with HIV survive (one-tailed RI p-value = 0.108), 2.7 percentage points more likely to say that ARVs helped prevent the spread of HIV (one-tailed RI p-value = 0.033), and 2.5 percentage points more likely to have heard of ARVs (one-tailed RI p-value = 0.098). However, treated respondents were not significantly more likely to respond that HIV/AIDS cannot be transmitted to coworkers (coefficient = 0.1 percentage point, one-tailed RI p-value = 0.466).

Indexing across all questions in line with our pre-registration, screening attendees of the placebo drama answered 66% of questions correctly on average, whereas those who attended the *Wahapahapa* screening answered 68% (one-tailed RI p-value = 0.026). Taken together, the results indicate a small but statistically significant positive effect on knowledge about HIV transmission and treatment.

Table 3: Effect of HIV Drama on HIV Knowledge
18-Month Endline

	Index	Transmit	ARV Survive	ARV No Spread	ARV Prompted
Treatment Effect	0.022	0.001	0.03	0.027	0.025
Standard Error	(0.01)	(0.012)	(0.023)	(0.014)	(0.023)
RI p-value	0.026	0.466	0.108	0.033	0.098
Test Direction	(+)	(+)	(+)	(+)	(+)
Effect Size (d)	0.14	0.01	0.08	0.14	0.07
Placebo Mean	0.66	0.96	0.83	0.04	0.83
Placebo Village SD	0.03	0.03	0.07	0.04	0.07
DV Range	[1-0]	[1-0]	[1-0]	[1-0]	[1-0]
Village Variance	0.000	0.000	0.000	0.000	0.000
Residual Variance	0.026	0.035	0.135	0.051	0.131
ICC (Village)	0.001	0.000	0.000	0.000	0.000
df (Kenward-Roger)	13.9	13.8	13.9	13.9	13.8
Blocked FE	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No
Observations	1,018	1,010	1,018	1,018	1,012
Villages	30	30	30	30	30

Note: Estimates and standard errors are from linear multilevel mixed models estimated by restricted maximum likelihood, with randomization-block fixed effects and a village random intercept; standard errors and denominator degrees of freedom (df) use the Kenward-Roger small-sample correction. One-tailed randomization inference (RI) p-values are calculated with 2,000 permutations of village-level treatment assignment within matched ward pairs, testing the sharp null hypothesis of no effect for any observation. Village Variance, Residual Variance, and ICC (Village) report the estimated random-effect variance components and the conditional village-level intraclass correlation. Effect Size (d) divides the treatment effect by the placebo group's individual-level standard deviation, with Hedges' small-sample correction (the posttest-only variant of Morris 2008); it is distinct from Placebo Village SD, which is the standard deviation across placebo village means. Confidence intervals for all effect sizes appear in Appendix K. Full item text, response scale, and coding for each measure appear in Table A20.

4.2.2 Willingness to Disclose

The next group of outcomes assesses how the screening affected respondents' attitudes towards disclosure of a person's HIV status, the behavior that is most directly modeled in the radio drama.

Table 4 shows that respondents who attended the *Wahapahapa* screening were 7.4 percentage points more likely to say that they would disclose their status to a coworker if they tested positive for HIV (one-tailed RI p-value = 0.045) and 7.8 percentage points more likely to say that they would disclose their status to a friend (one-tailed RI p-value = 0.038). Treated respondents were also 6.9 percentage points more likely to say that if a family member is diagnosed with HIV, they should not keep it a secret (one-tailed RI p-value = 0.040).

Averaged across the three questions, respondents were 7.2 percentage points more likely to express support for HIV status disclosure than the placebo group, providing strong evidence of a positive treatment effect (one-tailed RI p-value = 0.034). The change is substantively significant, amounting to roughly two-thirds of a village-level standard deviation. Reluctance remained considerable: only 58% of placebo group respondents said they would tell a friend if they tested positive for HIV, compared with 65% of respondents in treatment villages.

Table 4: Effect of HIV Drama on HIV Disclosure, 18-Month Endline

	Index	No Secret	Coworker	Friend
Treatment Effect	0.072	0.069	0.074	0.078
Standard Error	(0.034)	(0.036)	(0.039)	(0.04)
RI p-value	0.034	0.04	0.045	0.038
Test Direction	(+)	(+)	(+)	(+)
Effect Size (d)	0.17	0.14	0.15	0.16
Placebo Mean	0.57	0.64	0.49	0.58
Placebo Village SD	0.11	0.1	0.13	0.13
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Village Variance	0.004	0.003	0.004	0.005
Residual Variance	0.155	0.215	0.242	0.230
ICC (Village)	0.026	0.014	0.018	0.021
df (Kenward-Roger)	14	13.9	13.9	13.9
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Observations	1,018	1,018	1,015	1,015
Villages	30	30	30	30

Note: Exact one-tailed RI p-values from 2,000 permutations; estimation and inference as in Table 3: multilevel mixed models with randomization-block fixed effects and village random intercepts. Full item text and coding appear in Table A20.

4.2.3 Expressed and Perceived Stigma

We did not observe significant treatment effects of the HIV drama on expressed stigma towards HIV-positive persons at endline. Respondents who attended the *Wahapahapa* screening were slightly more

likely to be willing to sit next to someone with HIV/AIDS on the bus, and slightly more likely to say that an HIV-positive person should not be fired from their job. When asked if they believed others in their community would accept a person with HIV in the workplace, respondents in the treatment group were 2.5 percentage points (one-tailed RI p-value = 0.142) more likely to indicate that others in the community believe that the worker should be allowed to keep their job. While these treatment effects move in the predicted direction, none rise to conventional levels of statistical significance. This is due at least in part to the low levels of expressed stigma across the sample in the first place: on a scale from 0 (all stigmatizing attitudes) to 1 (no stigmatizing attitudes), the average index score in the placebo group was 0.94.

Table 5: Effect of HIV Drama on HIV Stigma, 18-Month Endline

	Index	Sit on Bus	Not Fired	Not Fired Norm
Treatment Effect	0.01	0.003	0.003	0.025
Standard Error	(0.013)	(0.011)	(0.015)	(0.024)
RI p-value	0.206	0.356	0.41	0.142
Test Direction	(+)	(+)	(+)	(+)
Effect Size (d)	0.07	0.02	0.02	0.08
Placebo Mean	0.94	0.97	0.97	0.89
Placebo Village SD	0.03	0.03	0.04	0.06
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Village Variance	0.001	0.000	0.001	0.002
Residual Variance	0.022	0.029	0.028	0.090
ICC (Village)	0.025	0.000	0.033	0.018
df (Kenward-Roger)	14	13.8	14	13.9
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Observations	1,018	1,016	1,016	1,014
Villages	30	30	30	30

Note: Exact one-tailed RI p-values from 2,000 permutations; estimation and inference as in Table 3: multilevel mixed models with randomization-block fixed effects and village random intercepts. Full item text and coding appear in Table A20.

4.2.4 Community Priorities

Finally, we measured the impact of the HIV drama on preferences for local politicians and community priorities.

When enumerators presented respondents with a choice between two hypothetical candidates running for village chairperson, respondents in the placebo group were roughly equally likely to prefer a candidate running on a platform to increase access to HIV/AIDS services and a candidate running on a platform to improve roads or reduce crime. Respondents in the treatment group were 8.4 percentage points more

likely to say they would prefer the candidate with the HIV platform (one-tailed RI p-value = 0.011). To put this effect into perspective, this is larger than the experimentally assessed effect of changes in important features of the candidate's identity, such as gender and religious background.

We observe smaller treatment effects on respondents' ranking of HIV as a community priority. Given a set of nine cards representing common challenges in rural communities, 13% of placebo group respondents selected HIV/AIDS care as the most important issue facing their village. Treated respondents were 2.9 percentage points more likely to rank HIV/AIDS first, although the results do not rise to conventional levels of statistical significance (one-tailed RI p-value = 0.183).

When we average the two pre-registered measures of community prioritization, we observe a 5.5 percentage point effect of the *Wahapahapa* treatment (one-tailed RI p-value = 0.011), amounting to roughly three-quarters of a village-level standard deviation.

Table 6: Effect of HIV Drama on HIV Priorities, 18-Month Endline

	Index	Elect	HIV First
Treatment Effect	0.055	0.084	0.029
Standard Error	(0.02)	(0.031)	(0.027)
RI p-value	0.011	0.011	0.183
Test Direction	(+)	(+)	(+)
Effect Size (d)	0.17	0.17	0.09
Placebo Mean	0.33	0.51	0.13
Placebo Village SD	0.07	0.12	0.05
DV Range	[1-0]	[1-0]	[1-0]
Village Variance	0.000	0.000	0.002
Residual Variance	0.104	0.244	0.121
ICC (Village)	0.000	0.000	0.014
df (Kenward-Roger)	13.9	13.9	13.9
Blocked FE	Yes	Yes	Yes
Covariates	No	No	No
Observations	1,019	1,019	987
Villages	30	30	30

Note: Exact one-tailed RI p-values from 2,000 permutations; estimation and inference as in Table 3: multilevel mixed models with randomization-block fixed effects and village random intercepts. Full item text and coding appear in Table A20.

4.3 Comparing Effects at Two Weeks and 18 Months

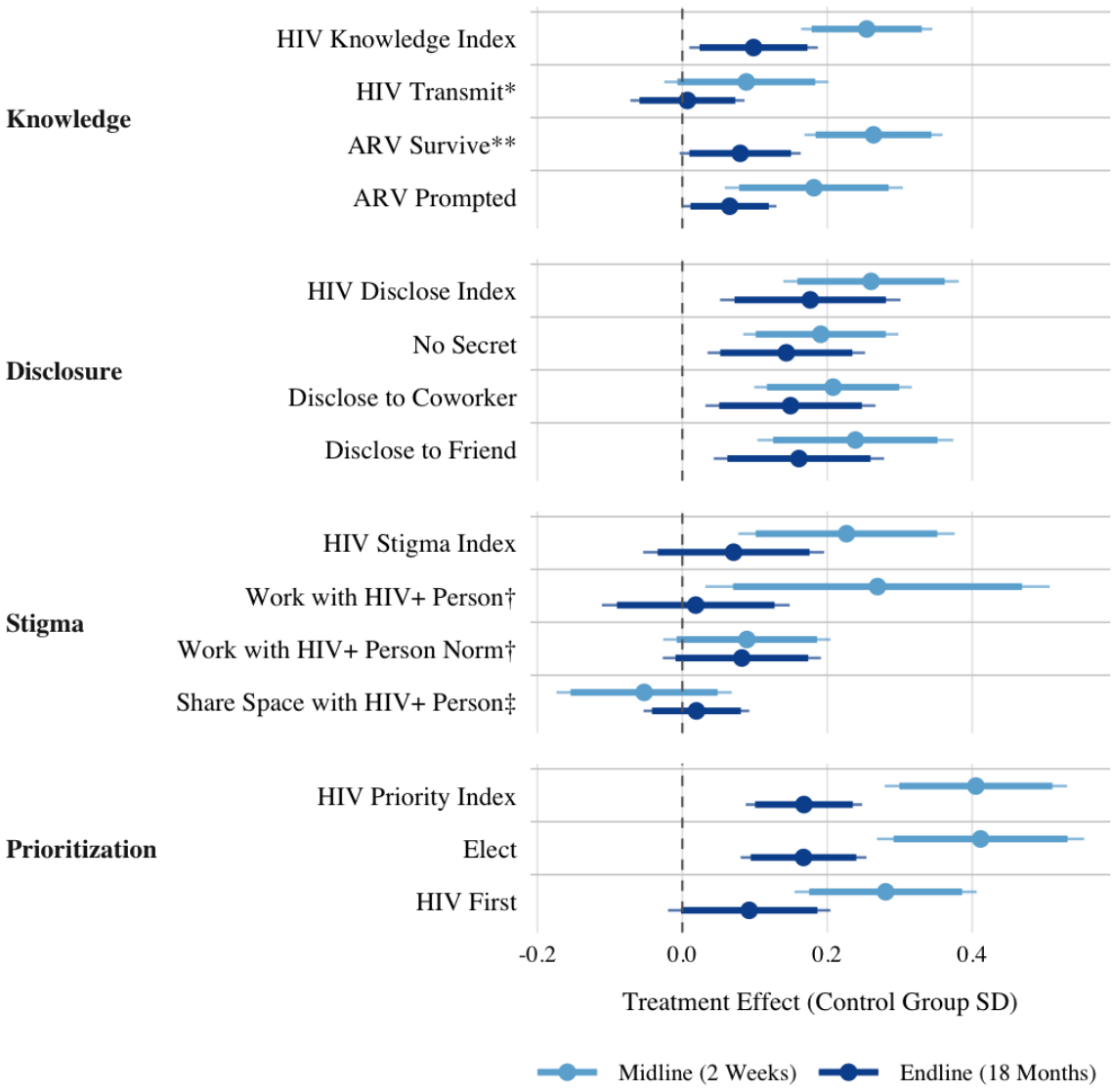
How do the results at 18 months compare to observed results two weeks after treatment? Figure 2 compares treatment effects at midline (two weeks post-intervention, originally reported in [REDACTED for review]) and endline (18 months post-intervention). While treatment effects remained detectable 18

months after treatment, we observed treatment effect decay across all outcome indices (full estimates in subsection D.1).

Figure A1 compares rates of decay across all outcome measures. Endline treatment effects on respondents' knowledge about HIV were 40% of the midline effect (midline coefficient = 0.25 standard deviations, endline coefficient = 0.10 standard deviations), 69% for respondents' willingness to disclose HIV status (midline coefficient = 0.26 standard deviations, endline coefficient = 0.18 standard deviations), 32% for respondents' expressed stigma (midline coefficient = 0.22 standard deviations, endline coefficient = 0.07 standard deviations), and 43% for respondents' HIV prioritization (midline coefficient = 0.4 standard deviations, endline coefficient = 0.17 standard deviations). Treatment effect decay is broadly consistent across indices and their components. On average, effects measured 18 months after treatment were about half the size of effects measured after two weeks.

How do these rates of decay compare to the broader entertainment-education literature? Almost all studies that report both short-term and long-term effects find some decay (Rahmani et al. 2025). In comparable radio-drama experiments in Tanzania, treatment effects on the rejection of forced marriage retained 27% of their midline magnitude after 15 months (Green et al. 2023), and effects on gender-based violence outcomes retained between 40% and 91% after 16–17 months (Montano et al. 2025; Montano 2026). The effect of a podcast intervention on Ugandan parents' assessment of health claims retained 43% of its immediate magnitude after one year (Semakula et al. 2020). Our retention rates fall in the middle of this range.

Figure 2: **Effect of HIV Drama on Key Outcomes: Midline v Endline**



Note: Thick lines show 90% confidence intervals; thin lines show 95% confidence intervals. All outcomes standardized by placebo group mean and SD within each time period. Outcomes marked with a superscript symbol asked slightly different questions at midline and endline. * = Midline: HIV transmission during pregnancy; Endline: HIV transmission in the workplace. ** = Midline: Remembering "special drugs" to help a person with HIV live longer; Endline: The most important things a person with HIV could do to live longer. † = Midline: Acceptance of HIV in the workplace; Endline: Whether a person with HIV should be fired. ‡ = Midline: Willingness to share a house with a person with HIV; Endline: Willingness to sit on a bus next to a person with HIV.

4.4 Spillover Results

In this section, we ask whether the treatment effects reported in the preceding section spilled over to the partners, friends, and children of treated respondents. We report estimated treatment effects on indices

of knowledge, disclosure, stigma, and prioritization outcomes. Treatment effect estimates on individual index items are shown in Appendix subsection D.2.

4.4.1 Spillover Effects on Partners

Table 7 shows the estimated effects of the *Wahapahapa* treatment on the partners of attendees of the HIV drama. Partners of treated respondents were not discernibly more knowledgeable about HIV or less likely to stigmatize HIV-positive persons than their counterparts in the placebo group. Estimated treatment effects are below 1 percentage point and statistically insignificant for both outcomes, consistent with the small long-term treatment effects on screening attendees.

On the other hand, treatment effects on attitudes towards HIV status disclosure appear to spill over to screening attendees' partners. Partners in treatment villages scored 6.6 percentage points higher than partners in placebo villages on the HIV disclosure index, almost equal to the treatment effect on the respondents (one-tailed RI p-value = 0.042).

We also observe evidence that changes in political priorities diffused to attendees' partners. Partners in treatment villages scored 4.8 percentage points higher on an index of political priorities combining support for a hypothetical candidate advocating for HIV/AIDS health care and rank-ordering of HIV/AIDS as a community issue. The effect amounts to roughly 0.7 village-level standard deviations, although it falls just short of conventional standards for statistical significance (one-tailed RI p-value = 0.078).

Table 7: Spillover Effects on Partners

	Knowledge	**Disclosure**	**Stigma**	**Priority**
	Index	Index	Index	Index
Treatment Effect	0.009	0.066	-0.005	0.048
Standard Error	(0.014)	(0.036)	(0.021)	(0.029)
RI p-value	0.278	0.042	0.577	0.078
Test Direction	(+)	(+)	(+)	(+)
Effect Size (d)	0.05	0.16	-0.02	0.16
Placebo Mean	0.64	0.54	0.9	0.32
Placebo Village SD	0.05	0.13	0.06	0.07
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Village Variance	0.000	0.003	0.001	0.002
Residual Variance	0.034	0.158	0.051	0.092
ICC (Village)	0.000	0.018	0.019	0.026
df (Kenward-Roger)	13.8	13.9	13.9	13.9
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Observations	709	709	709	709
Villages	30	30	30	30

Note: Exact one-tailed RI p-values from 2,000 permutations; estimation and inference as in Table 3: multilevel mixed models with randomization-block fixed effects and village random intercepts. Full item text and coding appear in Table A20.

4.4.2 Spillover Effects on Friends

We find no evidence of positive spillover effects to the friends of those who attended the *Wahapahapa* screening. No estimated treatment effects in the predicted direction exceed 1.5 percentage points, and no effects are statistically significant at conventional levels. There is some evidence that respondents' friends became *more* likely to express stigma towards HIV-positive individuals (coefficient = -6.4pp, one-tailed RI p-value = 0.986), although we interpret this finding cautiously since it runs against the predicted effect direction and there was no effect on the main screening attendees.

Table 8: Spillover Effects on Friends

	Knowledge	**Disclosure**	**Stigma**	**Priority**
	Index	Index	Sit on Bus	Index
Treatment Effect	0.003	0.01	-0.064	0.014
Standard Error	(0.039)	(0.032)	(0.026)	(0.024)
RI p-value	0.482	0.368	0.986	0.286
Test Direction	(+)	(+)	(+)	(+)
Effect Size (d)	0.01	0.02	-0.20	0.04
Placebo Mean	0.72	0.57	0.88	0.36
Placebo Village SD	0.12	0.09	0.07	0.07
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Village Variance	0.008	0.001	0.001	0.001
Residual Variance	0.091	0.212	0.123	0.103
ICC (Village)	0.084	0.005	0.011	0.010
df (Kenward-Roger)	14	13.8	13.8	13.8
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Observations	972	972	972	972
Villages	30	30	30	30

Note: Exact one-tailed RI p-values from 2,000 permutations; estimation and inference as in Table 3: multilevel mixed models with randomization-block fixed effects and village random intercepts. Full item text and coding appear in Table A20.

4.4.3 Spillover Effects on Teenage Children

Finally, diffusion of treatment effects to teenage children is limited to attitudes towards HIV status disclosure. Children of treated respondents were 11.7 percentage points more likely to support HIV status disclosure within the family, equal to a full village-level standard deviation and the largest effect in the study (one-tailed RI p-value = 0.016). Evidently, discussion within the family about disclosing HIV status reached the teenagers of the household, shifting within-family norms about disclosure. Treatment effects on the remaining outcomes measured for children (HIV knowledge, stigma, and the hypothetical election) are small and statistically indistinguishable from zero (Table 9).

Table 9: Spillover Effects on Teenage Children

	Knowledge	**Disclosure**	**Stigma**		**Priority**
	ARV Open-Ended	Family Keep Secret	Sit on Bus	Parents Advice for Bus	Elect
Treatment Effect	-0.012	0.117	-0.007	-0.011	-0.018
Standard Error	(0.039)	(0.05)	(0.06)	(0.054)	(0.044)
RI p-value	0.608	0.016	0.553	0.581	0.701
Test Direction	(+)	(+)	(+)	(+)	(+)
Effect Size (d)	-0.03	0.23	-0.01	-0.02	-0.04
Placebo Mean	0.85	0.52	0.67	0.66	0.67
Placebo Village SD	0.11	0.12	0.18	0.18	0.13
DV Range	[1-0]	[1-0]	[1-0]	[1-0]	[1-0]
Village Variance	0.002	0.003	0.011	0.006	0.000
Household Variance	0.012	0.000	0.028	0.034	0.000
Residual Variance	0.118	0.234	0.183	0.182	0.224
ICC (Village)	0.016	0.012	0.051	0.027	0.000
df (Kenward-Roger)	13.1	13	13.5	13.3	12.6
Blocked FE	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No
Observations	496	495	496	495	496
Villages	30	30	30	30	30

Note: Exact one-tailed RI p-values from 2,000 permutations; estimation and inference as in Table 3, with an additional household random intercept nested within village (Household Variance), as some households contribute more than one child. Full item text and coding appear in Table A20.

4.5 Harms

Qualitative reports documenting participant reactions and post-screening discussions were collected in both treatment and placebo villages throughout the intervention period. These reports revealed no adverse participant reactions to *Wahapahapa* or *Tamapendo*, including stigma-promoting effects, respondent distress, or other unintended harms.

5 Discussion

5.1 Summary of Findings

Eighteen months after a single 1.5-hour screening, attendees of the HIV drama *Wahapahapa* remained more knowledgeable about HIV treatment, more supportive of HIV status disclosure, and more likely to prioritize HIV policy. Disclosure attitudes diffused to attendees' partners and teenage children, and policy prioritization showed signs of diffusion to partners. On the other hand, HIV knowledge and expressed stigma did not diffuse to any group, and no effects diffused to friends. Qualitative monitoring throughout the intervention period detected no adverse reactions, respondent distress, or stigma-promoting effects in either arm (subsection 4.5).

5.2 Interpretation

Treatment effects measured 18 months after the intervention decayed to roughly half the magnitude of effects measured 2 weeks after treatment, consistent with comparable studies (Montano et al. 2025; Green et al. 2023). A single screening can durably shift attitudes, but repeated exposure may be needed to maintain the full magnitude of initial effects.

Why did the drama's effects persist for some outcomes but not others? The decay in treatment effects on factual knowledge is consistent with prior findings that information acquired from a single exposure fades without reinforcement (Gerber et al. 2011; Kalla and Broockman 2021). Expressed stigma was already low in the placebo group, so there was little room for positive change. Attitudes toward disclosure and policy prioritization are normative and evaluative positions rather than discrete facts, and both concern themes that *Wahapahapa* dramatized at length through its protagonist's disclosure decisions. Such positions may be sustained by ongoing household conversation, an interpretation reinforced by the spillover results: precisely the disclosure and prioritization outcomes, and not knowledge, diffused to partners and children.

Diffusion of treatment effects was selective. There is evidence that the drama's themes entered household conversations but did not diffuse to friends, suggesting that entertainment-education's "multiplier effects" operate through intimate relationships rather than broader social diffusion, at least with respect to HIV status disclosure (Valente et al. 1994; Chatterjee et al. 2009). Within-family transmission is especially consequential for HIV, where intra-family communication supports treatment uptake and the prevention of mother-to-child transmission (Betancourt et al. 2010; Rochat et al. 2011). One plausible explanation is that HIV remains a topic people discuss with spouses and children but not with friends.

5.3 Limitations

There are important limitations to the present study. First, our outcomes are self-reported attitudes and behavioral intentions rather than observed behaviors or health outcomes. Second, with 30 clusters, some of our null results may reflect insufficient power rather than genuinely zero effects. Appendix I reports minimum detectable effects for each outcome. Third, our design cannot distinguish between direct conversation about the drama's content and more diffuse shifts in household climate around HIV. Future studies could disentangle these mechanisms by collecting more granular data on household communication. Fourth, although surveyed friends were drawn from each respondent's three closest friendships

(which implies some regularity of contact), we did not directly measure the frequency or intimacy of contact between respondents and their nominated friends, and we did not conduct qualitative interviews. We therefore cannot rule out infrequent contact as an explanation for the absence of friend spillovers. We do find that many respondents are unwilling to talk about HIV with friends (only 58% of placebo group respondents and 65% of treatment group respondents would disclose a positive diagnosis to a friend), so even close friendships may not support conversations about the drama's themes.

5.4 Generalisability

Our estimates are representative of adults aged 18 to 65 in the study villages; primary analyses focus on screening attendees, with full-sample intention-to-treat estimates in subsection D.4. Our results therefore may not generalize to older residents or children. Moreover, the study villages are rural, relatively remote communities in northeastern Tanzania. Effects may therefore differ in urban settings or where media environments are denser. The intervention was delivered at communal listening events rather than through individual home listening, so more research is needed into the effects of narrative media in more naturalistic settings.

5.5 Conclusion

Entertainment-education can cause changes in HIV-related attitudes that persist at least 18 months after a single exposure and diffuse through household relationships. These findings suggest productive directions for future work. Studies pairing naturalistic entertainment-education interventions with behavioral data on clinic visits, testing rates, and antiretroviral therapy enrollment would clarify whether attitudinal shifts translate into behavioral change. Research incorporating more granular information about household communication would help clarify how narrative content enters family conversations. Most importantly, the basic research design requires replication in new settings and on new topics to solidify our understanding of whether and how entertainment-education effects persist over time and diffuse through familial and social networks.

References

Appel, M. and Richter, T. (2007). Persuasive effects of fictional narratives increase over time. *Media Psychology*, 10(1):113–134.

Audi, C., Jahanpour, O., Antelman, G., Guay, L., Rutaihwa, M., van de Ven, R., Woelk, G., and Baird, S. J. (2021). Facilitators and barriers to antiretroviral therapy adherence among HIV-positive adolescents living in Tanzania. *BMC Public Health*, 21(1):2274.

Avert (2020). HIV and AIDS in East and Southern Africa regional overview. Technical report.

Bandura, A. (2004). Health Promotion by Social Cognitive Means. *Health Education & Behavior*, 31(2):143–164.

Banerjee, A., Barnhardt, S., and Duflo, E. (2015). Movies, margins, and marketing: Encouraging the adoption of iron-fortified salt. In *Insights in the economics of aging*, pages 285–306. University of Chicago Press.

Banerjee, A., Ferrara, E. L., and Orozco-Olvera, V. H. (2019). The Entertaining Way to Behavioral Change: Fighting HIV with MTV. Technical Report w26096, National Bureau of Economic Research.

Basaran, A.-M. B., Christensen, J. L., Miller, L. C., Appleby, P. R., and Read, S. J. (2019). The relationship between social norms and sexual risk-reduction intentions and behavior among men who have sex with men: Findings from an eHealth intervention. *Psychology of addictive behaviors : journal of the Society of Psychologists in Addictive Behaviors*, 33(4):382–391.

Betancourt, T. S., Abrams, E. J., McBain, R., and Fawzi, M. C. S. (2010). Family-centred approaches to the prevention of mother to child transmission of HIV. *Journal of the International AIDS Society*, 13(S2):S2–S2. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1186/1758-2652-13-S2-S2.

Bjorvatn, K., Cappelen, A. W., Sekei, L. H., Sørensen, E. Ø., and Tungodden, B. (2020). Teaching through television: experimental evidence on entrepreneurship education in Tanzania. *Management Science*, 66(6):2308–2325.

Boulay, M., Storey, J. D., and Sood, S. (2002). Indirect Exposure to a Family Planning

Mass Media Campaign in Nepal. *Journal of Health Communication*, 7(5):379–399. _eprint: https://doi.org/10.1080/10810730290001774.

Cacioppo, J. T., Petty, R. E., Kao, C. F., and Rodriguez, R. (1986). Central and peripheral routes to persuasion: An individual difference perspective. *Journal of Personality and Social Psychology*, 51(5):1032–1043.

Carey, M. P., Senn, T. E., Walsh, J. L., Coury-Doniger, P., Urban, M. A., Fortune, T., Vanable, P. A., and Scott-Sheldon, L. A. J. (2015). Evaluating a brief, video-based sexual risk reduction intervention and assessment reactivity with STI clinic patients: results from a randomized controlled trial. *AIDS and Behavior*, 19(7):1228–1246.

Centola, D. and Macy, M. (2007). Complex Contagions and the Weakness of Long Ties. *American Journal of Sociology*, 113(3):702–734.

Chatterjee, J. S., Bhanot, A., Frank, L. B., Murphy, S. T., and Power, G. (2009). The Importance of Interpersonal Discussion and Self-Efficacy in Knowledge, Attitude, and Practice Models. *International Journal of Communication*, 3:28–28.

Dahlstrom, M. F. (2012). The Persuasive Influence of Narrative Causality: Psychological Mechanism, Strength in Overcoming Resistance, and Persistence Over Time. *Media Psychology*, 15(3):303–326.

Doré, A., Rebitschek, F. G., Morello, K., Stierstädter, E., Thiemann, K. J., and Kubiak, T. (2025). A systematic review and meta-analysis of the effects of Entertainment-Education interventions on persuasive health outcomes. *Scientific Reports*, 15(1):27939.

Fonner, V. A., Mbwambo, J., Kennedy, C. E., Kerrigan, D., and Sweat, M. D. (2020). Do sexual partners talk to each other about HIV? Exploring factors associated with HIV-related partner communication among men and women in Tanzania. *AIDS and Behavior*, 24:891–902.

Frank, L. B. and Falzone, P., editors (2021). *Entertainment-Education Behind the Scenes: Case Studies for Theory and Practice*. Palgrave Macmillan, Cham.

Gerber, A. S., Gimpel, J. G., Green, D. P., and Shaw, D. R. (2011). How Large and Long-lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment. *American Political Science Review*, 105(1):135–150.

Green, D. P., Groves, D. W., Manda, C., Montano, B., and Rahmani, B. (2023). A Radio Drama's Effects on Attitudes Toward Early and Forced Marriage: Results From a Field Experiment in Rural Tanzania. *Comparative Political Studies*, 56(8):1115–1155.

Green, D. P., Groves, D. W., Manda, C., Montano, B., and Rahmani, B. (2024). The Effects of Independent Local Radio on Tanzanian Public Opinion: Evidence from a Planned Natural Experiment. *The Journal of Politics*, 86(1):231–240.

Green, D. P., Wilke, A. M., and Cooper, J. (2020). Countering Violence Against Women by Encouraging Disclosure: A Mass Media Experiment in Rural Uganda. *Comparative Political Studies*, 53(14):2283–2320.

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.

Hu, Y., Li, Q., and Chen, Y. (2018). Evaluation of two health education interventions to improve the varicella vaccination: a randomized controlled trial from a province in the east China. *BMC Public Health*, 18(1):144.

Jones, C. W. and Paris, C. (2018). It's the End of the World and They Know It: How Dystopian Fiction Shapes Political Attitudes. *Perspectives on Politics*, 16(4):969–989.

Kalichman, S. C., Simbayi, L. C., Jooste, S., Toefy, Y., Cain, D., Cherry, C., and Kagee, A. (2005). Development of a Brief Scale to Measure AIDS-Related Stigma in South Africa. *AIDS and Behavior*, 9(2):135–143.

Kalla, J. and Broockman, D. (2021). "Outside Lobbying" over the Airwaves: A Randomized Field Experiment on Televised Issue Ads. *American Political Science Review*, 116:1–7.

Katz, E. (1957). The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis*. *Public Opinion Quarterly*, 21(1):61–78.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, 75(1):83–119.

Mdee, A. and Mushi, A. (2021). Untangling blame and responsibility for service delivery and local governance performance: testing a grounded social accountability approach in Tanzania. *Local Government Studies*, 47(6):993–1013.

Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC) [Tanzania Mainland], Ministry of Health (MoH) [Zanzibar], National Bureau of Statistics (NBS), Office of the Chief Government Statistician (OCGS), and ICF (2016). Tanzania Demographic and Health Survey and Malaria Indicator Survey (TDHS-MIS) 2015–16. Technical report, MoHCDGEC, MoH, NBS, OCGS, and ICF, Dar es Salaam, Tanzania, and Rockville, Maryland, USA.

Mohammed, S. (2001). Personal Communication Networks and the Effects of an Entertainment-Education Radio Soap Opera in Tanzania. *Journal of Health Communication*, 6(2):137–154. _eprint: https://doi.org/10.1080/10810730117219.

Montano, B. (2026). *The Politics of Gender Norms: Evidence from Rural Tanzania*. PhD thesis, Columbia University.

Montano, B., Emmanuel, S., Green, D. P., Groves, D. W., and Rahmani, B. (2025). Increasing Awareness of Gender-Based Violence Outside of the Household: Experimental Evidence from a Field Experiment in Rural Tanzania. *Working Paper*.

Murphy, E. (2004). Diffusion of Innovations: Family Planning in Developing Countries. *Journal of Health Communication*, 9(sup1):123–129. _eprint: https://doi.org/10.1080/10810730490271566.

National Ministry of Health, Community Development, Gender, Elderly, and Children (2017). National survey on the drivers and consequences of child marriage in Tanzania. Technical report.

Paluck, E. L. and Green, D. P. (2009). Deference, Dissent, and Dispute Resolution: An Experimental Intervention Using Mass Media to Change Norms and Behavior in Rwanda. *American Political Science Review*, 103(4):622–644.

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*, pages 1–23.

Rochat, T. J., Bland, R., Coovadia, H., Stein, A., and Newell, M.-L. (2011). Towards a Family-Centered Approach to HIV Treatment and Care for HIV-Exposed Children, Their Mothers and Their Families in Poorly Resourced Settings. *Future Virology*, 6(6):687–696. _eprint: https://doi.org/10.2217/fvl.11.45.

Rogers, E. M., Vaughan, P. W., Swalehe, R. M., Rao, N., Svenkerud, P., and Sood, S. (1999). Effects of an Entertainment-education Radio Soap Opera on Family Planning Behavior in Tanzania. *Studies in Family Planning*, 30(3):193–211. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1728-4465.1999.00193.x.

Sanga, E. S., Mukumbang, F. C., Mushi, A. K., Lerebo, W., and Zarowsky, C. (2019). Understanding factors influencing linkage to HIV care in a rural setting, Mbeya, Tanzania: qualitative findings of a mixed methods study. *BMC Public Health*, 19(1):383.

Semakula, D., Nsangi, A., Oxman, A. D., Oxman, M., Austvoll-Dahlgren, A., Rosenbaum, S., Morelli, A., Glenton, C., Lewin, S., Nyirazinyoye, L., Kaseje, M., Chalmers, I., Fretheim, A., Rose, C. J., and Sewankambo, N. K. (2020). Effects of the Informed Health Choices podcast on the ability of parents of primary school children in Uganda to assess the trustworthiness of claims about treatment effects: one-year follow up of a randomised trial. *Trials*, 21:187.

Sharan, M. and Valente, T. W. (2002). Spousal Communication and Family Planning Adoption: Effects of a Radio Drama Serial in Nepal. *International Family Planning Perspectives*, 28(1):16–25.

Slater, M. D. and Rouner, D. (2002). Entertainment—Education and Elaboration Likelihood: Understanding the Processing of Narrative Persuasion. *Communication Theory*, 12(2):173–191. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-2885.2002.tb00265.x.

Southwell, B. and Yzer, M. C. (2007). The Roles of Interpersonal Communication in Mass Media Campaigns. *Communication Yearbook*, 31(1):420–462.

Southwell, B. G. and Yzer, M. C. (2009). When (and Why) Interpersonal Talk Matters for Campaigns. *Communication Theory*, 19(1):1–8.

UNAIDS (2022). Global AIDS Monitoring 2023: Indicators and questions for monitoring progress on the 2021 Political Declaration on HIV and AIDS. Technical report, Joint United Nations Programme on HIV/AIDS, Geneva.

Valente, T., Kim, Y., Lettenmaier, C., Glass, W., and Dibba, Y. (1994). Radio Promotion of Family Planning in The Gambia. *International Family Planning Perspectives*, 20:96.

Vaughan, P. W. and Rogers, E. M. (2000). A staged model of communication effects: Evidence from an entertainment–education radio soap opera in Tanzania. *Journal of Health Communication*, 5(3):203–227. Place: United Kingdom.

Wamoyi, J., Fenwick, A., Urassa, M., Zaba, B., and Stones, W. (2010). Parent-child communication about sexual and reproductive health in rural Tanzania: Implications for young people's sexual health interventions. *Reproductive Health*, 7:6.

Wilke, A. M., Green, D. P., and Cooper, J. (2020). A Placebo Design to Detect Spillovers from an Education–Entertainment Experiment in Uganda. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 183(3):1075–1096.

Online Appendix for

The Long-Term and Spillover Effects of a Radio Drama on HIV Knowledge, Expressed Stigma, Willingness to Disclose, and Political Priorities: Evidence from a Cluster-Randomized Trial in Tanzania

August 10, 2026

*The authors would like to thank the implementing partner, Media for Development International, led by John Ribber. We are grateful to the entire research team at Innovations for Poverty Action, particularly Martin Zuakulu, Rachel Jones, Kyu Khin Gar, Zachary Isdahl, Robert Mwandumbya, Neema Msechu, Gilbert Loshoock, Cosmas Sway, Zinabu Omary, Fatuma Yahaya, Frank Simon, Fadhili Mashaka, Stephano Kizigo, and Saulo Mwakisongo. The studies described in this paper were reviewed and approved by Columbia's Institutional Review Board (protocol AAAR5582) and Tanzania's Commission of Science and Technology (protocol 14528). The trial was preregistered with EGAP on the Open Science Framework (https://osf.io/4tdz2, registered August 10, 2019); the pre-analysis addendum covering the spillover surveys, finalized before spillover data collection began, is available in the study's OSF repository (https://osf.io/c8tfu, posted January 28, 2021) and included in the appendix. These studies were made possible by financial support from Wellspring Philanthropic Fund. Wellspring Philanthropic Fund had no role in the study design, conduct, analysis, or reporting. Corresponding author: Dylan W. Groves (dylanwgroves@gmail.com).

Appendix Contents

A CONSORT Checklist for Cluster Randomised Trials

The table below maps each item of the CONSORT 2010 checklist, as extended for cluster randomised trials, to the location in the manuscript or appendix where the item is reported. Item descriptions are given in condensed form; cluster-specific requirements of the extension are incorporated into the relevant item descriptions.

Table A1: CONSORT 2010 Checklist, Cluster Trial Extension

Item	Description	Reported in
Title and Abstract		
1a	Identification as a cluster randomised trial in the title	Title page; Abstract
1b	Structured summary of trial design, methods, results, and conclusions	Abstract
Introduction		
2a	Scientific background and explanation of rationale	Introduction; Literature Review & Theoretical Background
2b	Specific objectives or hypotheses, and how they apply to clusters and to individual participants	Literature Review & Theoretical Background, "The Present Study: Aims and Hypotheses"
Methods: Trial Design		
3a	Description of trial design, including definition of the cluster, the level at which the intervention was allocated, and allocation ratio	Methods, "Sample Selection & Randomization" (parallel-group design; cluster = village; 1:1 allocation within matched ward pairs)
3b	Important changes to methods after trial commencement, with reasons	Methods, "Statistical Analyses" (outcome measures for the endline and spillover surveys added in a registered update prior to endline data collection; no other changes)
Methods: Participants		
4a	Eligibility criteria for clusters and for individual participants	Methods, "Site Selection" and "Respondent Sampling"; full criteria in subsection E.3
4b	Settings and locations where the data were collected	Methods, "Study setting" and "Data Collection"
4c	Whether individual participants were identified and recruited before or after cluster randomisation	Methods, "Random Assignment" (all invited participants were recruited and enrolled before allocation was revealed to field teams)
Methods: Interventions		
5	The interventions for each group with sufficient detail to allow replication, including whether they were applied to clusters or to individual participants	Methods, "Intervention"; subsection E.2 (both dramas delivered once at village-level community listening events)
Methods: Outcomes		

Continued on next page

Table A1 continued from previous page

Item	Description	Reported in
6a	Completely defined pre-specified primary and secondary outcome measures, including whether measured at the cluster or individual level	Methods, "Outcome Measures" and "Statistical Analyses"; Table A20; Table A21 (all outcomes measured at the individual level)
6b	Any changes to trial outcomes after the trial commenced, with reasons	Methods, "Statistical Analyses" (endline and spillover outcomes registered before endline data collection)
Methods: Sample Size		
7a	How sample size was determined, including the number of clusters, cluster size, and the intracluster correlation assumed	Methods, "Sample Size" (30 clusters of 40 respondents, set by available funding together with a prospective power assessment targeting minimum detectable effects of roughly 0.2 SD; intracluster correlation unknown ex ante); realized minimum detectable effects and unconditional intracluster correlations in Appendix I; intracluster correlations conditional on the matched-pair blocking in Appendix M
7b	When applicable, explanation of any interim analyses and stopping guidelines	Methods, "Data Collection" (no interim analyses or stopping guidelines were planned or conducted)
Methods: Randomisation		
8a	Method used to generate the random allocation sequence for clusters	Methods, "Random Assignment" (random assignment performed in Stata; code in the replication materials)
8b	Type of randomisation and details of any restriction	Methods, "Random Assignment" (restricted randomisation: one village per matched ward pair)
9	Mechanism used to implement the allocation sequence and steps taken to conceal the sequence until interventions were assigned	Methods, "Random Assignment" (assigned drama carried in a sealed envelope, opened only at the start of the listening event)
10	Who generated the allocation sequence, who enrolled clusters, and who assigned clusters to interventions	Methods, "Random Assignment" (sequence generated by a research team member separate from the enumerators who enrolled participants)
11a	Blinding: if done, who was blinded after assignment to interventions and how	Methods, "Random Assignment" and "Data Collection" (participants and enumerators at every wave were blind to village treatment assignment)
11b	If relevant, description of the similarity of interventions	Methods, "Intervention" (the placebo drama was matched to the treatment on format, length, delivery mode, and setting, differing only in topic)
Methods: Statistical Methods		

Continued on next page

Table A1 continued from previous page

Item	Description	Reported in
12a	Statistical methods used to compare groups for primary and secondary outcomes, including how clustering was accounted for	Methods, "Statistical Analyses" (linear mixed model with randomization-block fixed effects and a village random intercept, estimated by REML with Kenward-Roger small-sample corrections; hypothesis testing by randomization inference at the level of the randomized unit; separate analyses by respondent sample; design-based estimates with village-clustered standard errors in Appendix L)
12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	Methods, "Statistical Analyses"; Appendix L; Appendix J; Appendix M; Appendix H; Appendix K; subsection D.4
Results: Participant Flow		
13a	For each group, the numbers of clusters and of individual participants randomly assigned, receiving the intended treatment, and analysed for the primary outcome	Figure 1; Results, "Preliminary Results: Attrition and Compliance"; per-group N in each results table
13b	For each group, losses and exclusions after randomisation, with reasons	Results, "Preliminary Results: Attrition and Compliance"; Table 2; Appendix G
Results: Recruitment		
14a	Dates defining the periods of recruitment and follow-up	Methods, "Data Collection" (baseline and screenings July–September 2019; midline August–September 2019; endline December 2020–March 2021; friend and child surveys January–March 2021)
14b	Why the trial ended or was stopped	Methods, "Data Collection" (all planned waves were completed; the trial was not stopped early)
Results: Baseline Data		
15	A table showing baseline demographic and clinical characteristics for each group, at the cluster level and the individual participant level	Table A3 (cluster level); Table 1 (individual level, by analysis sample); Table A4 (full covariate balance)
Results: Numbers Analysed		
16	For each group, the number of clusters and participants included in each analysis, and whether the analysis was by original assigned groups	Figure 1; per-group N reported in every results table; intention-to-treat estimates for the full invited sample in subsection D.4
Results: Outcomes and Estimation		
17a	For each primary and secondary outcome, results for each group and the estimated effect size and its precision	Results, "Main Results" and "Spillover Results"; standardized effect sizes with 95% confidence intervals in Appendix K

Continued on next page

Table A1 continued from previous page

Item	Description	Reported in
17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	Results tables report absolute (percentage-point) effects with placebo-group means; standardized effect sizes in Appendix K
17c	Coefficient of intracluster correlation for each primary outcome	Reported alongside the treatment effect in every main-text results table; Appendix M
Results: Ancillary Analyses		
18	Results of any other analyses performed, distinguishing pre-specified from exploratory	Appendix H (exploratory); Appendix J; Appendix M; subsection D.4 and subsection D.1 (pre-specified)
Results: Harms		
19	All important harms or unintended effects in each group	Results, "Harms" (subsection 4.5); monitoring procedure described in Methods, "Intervention"
Discussion		
20	Trial limitations, addressing sources of potential bias, imprecision, and multiplicity of analyses	Discussion, "Limitations"; Appendix I
21	Generalisability (external validity, applicability) of the trial findings, to clusters and to individual participants	Discussion, "Generalisability"
22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	Discussion, "Summary of Findings", "Interpretation", and "Conclusion"
Other Information		
23	Registration number and name of trial registry	Methods, "Statistical Analyses" (registered on the Open Science Framework; identifier withheld for anonymous review)
24	Where the full trial protocol can be accessed, if available	subsection E.3
25	Sources of funding and other support, and role of funders	Acknowledgements (submitted as a separate file for anonymous review)

B TIDieR Checklist

The table below maps each item of the Template for Intervention Description and Replication (TIDieR) checklist to the location in the manuscript or appendix where the item is reported. Item descriptions are given in condensed form.

Table A2: TIDieR Checklist

Item	Description	Reported in
Brief Name		
1	Name or phrase describing the intervention	Methods, "Intervention" (110-minute abridged version of the Swahili radio drama *Wahapahapa*; placebo villages received the matched drama *Tamapendo*)
Why		
2	Rationale, theory, or goal of the intervention's essential elements	Introduction; Literature Review & Theoretical Background; subsection 2.3
What		
3	Materials used in the intervention, and where they can be accessed	Methods, "Intervention" (the two abridged audio dramas); subsection E.2 (content, storylines, and abridgment). The source dramas were produced by Media for Development International (*Wahapahapa*) and UZIKWASA (*Tamapendo*); full English translations of both abridged scripts and the audio recordings used in the trial are available from the authors on request.
4	Procedures, activities, and processes used, including enabling/support activities	Methods, "Intervention" (community listening events, full audio playback, attendance monitoring, refreshments, and no facilitated discussion)
Who Provided		
5	Expertise, background, and training of each category of intervention provider	Methods, "Intervention" (three-person teams of trained field officers; provider and monitoring roles described)
How		
6	Mode of delivery and whether individual or group	Methods, "Intervention" (single group audio playback at community listening events)
Where		
7	Type(s) of location(s), including necessary infrastructure	Methods, "Study setting," "Intervention," and "Site Selection" (community listening events in 30 rural villages across 15 wards in Tanga Region)

Continued on next page

Table A2 continued from previous page

Item	Description	Reported in
When and How Much		
8	Number/duration of sessions, schedule, dose	Methods, "Intervention" (one 110-minute session); Methods, "Data Collection" (screenings held July 25–September 10, 2019)
Tailoring		
9	Whether/how the intervention was personalized or adapted	No individual tailoring; each assigned drama was delivered uniformly and in full to all attendees within an experimental arm
Modifications		
10	Changes to the intervention during the study, with reasons	subsection E.2 (source drama condensed from the full ~120-hour series before fielding; no changes after trial commencement)
How Well		
11	Planned: how adherence/fidelity was assessed and by whom	Methods, "Intervention" (field officers monitored attendance and verified the correct drama was played in full)
12	Actual: extent to which the intervention was delivered as planned	Results, "Preliminary Results: Attrition and Compliance" (assigned drama played in full at every village, confirmed by field officers; 86.8% placebo-arm and 87.4% treatment-arm attendance, two-sided RI $p = 0.86$)

C Covariate Balance

Because treatment was assigned to villages rather than to individuals, we report baseline balance at both levels. Table A3 reports characteristics of the randomized units themselves; the tables that follow report characteristics of individual participants.

Table A3: **Cluster-Level Characteristics and Balance**

	Placebo	Treatment	RI p-value
Panel A: Village Characteristics			
Respondents enrolled	40.3 (0.6)	40.1 (0.3)	0.494
Distance to nearest town (km)	19.3 (10.0)	20.3 (11.5)	0.309
Distance to main road (km)	6.4 (2.5)	6.3 (3.8)	0.967
Distance to Pangani town (km)	40.9 (11.8)	41.4 (13.3)	0.615
Primary schools	1.23 (0.68)	1.07 (0.59)	0.556
Secondary school (1 = yes)	0.27 (0.46)	0.30 (0.46)	1.000
Health clinic (1 = yes)	0.27 (0.46)	0.53 (0.52)	0.296
Electricity (1 = yes)	0.67 (0.45)	0.63 (0.48)	1.000
Mosques	2.53 (1.61)	2.67 (1.68)	0.861
Churches	2.93 (2.90)	2.87 (2.01)	0.978
Villages	15	15	
Panel B: Village Means of Respondent Characteristics			
Age	37.43 (1.74)	38.40 (3.10)	0.330
Female	0.50 (0.00)	0.50 (0.00)	1.000
Reached Standard 7	0.75 (0.11)	0.77 (0.13)	0.733
Muslim	0.64 (0.28)	0.67 (0.25)	0.671
Owns a television	0.15 (0.09)	0.19 (0.11)	0.158
Owns a radio	0.44 (0.10)	0.46 (0.11)	0.405
Villages	15	15	

Note: The unit of analysis is the village ($N = 30$); cells report unweighted means across villages with standard deviations in parentheses. Panel A reports characteristics of the village itself: distances were measured during site selection, and school, clinic, electricity, and place-of-worship counts come from a survey of village leaders (two leaders per village, averaged; four villages have a single leader response). Panel B reports village means of baseline respondent characteristics; because respondents were sampled to be evenly split between men and women, the village mean for *Female* is fixed at 0.50 by design and carries no variation to test. Randomization p-values are two-sided and are calculated from the same 2,000 permutations of village-level treatment assignment within matched ward pairs used throughout the paper.

Table A4: **Balance Table: Treatment and Placebo Among Screening Attendees, All Baseline Covariates**

Variable	Placebo (N=524)	Treatment (N=525)	RI p-value
Gender (1 = Female)	0.52 (0.50)	0.53 (0.50)	0.792
Age (years)	38.23 (12.00)	38.71 (11.98)	0.648
Married (1 = Married)	0.65 (0.48)	0.65 (0.48)	0.952
Religion: Muslim (1 = Muslim)	0.64 (0.48)	0.65 (0.48)	0.861
Religiosity (services/week)	5.06 (10.62)	4.72 (10.15)	0.759
Head of Household	0.57 (0.50)	0.54 (0.50)	0.588
Household Size	4.85 (2.38)	5.03 (2.24)	0.431
Household Members Aged 0–18	2.56 (1.68)	2.70 (1.80)	0.379
Children Ever Born	3.51 (2.83)	3.65 (2.86)	0.432
Education: Standard 7+	0.76 (0.43)	0.76 (0.42)	0.891
Literate	0.82 (0.38)	0.83 (0.38)	0.748
Primary Language: Swahili	0.81 (0.39)	0.81 (0.39)	0.936
Years Lived in Village	23.73 (15.39)	25.47 (14.75)	0.222
Lived in Village at Age 16	0.55 (0.50)	0.58 (0.49)	0.675
Village Knowledge (1–4)	2.72 (0.76)	2.67 (0.75)	0.477
Ever Lived in City/Town	0.45 (0.50)	0.45 (0.50)	0.984
Never Visits Big City	0.16 (0.37)	0.18 (0.39)	0.391
Never Visits Town	0.31 (0.46)	0.37 (0.48)	0.148
Radio Consumption (0–5)	1.32 (1.61)	1.22 (1.48)	0.146
Any Radio, Last 3 Months	0.64 (0.48)	0.66 (0.48)	0.497
Ever Listens to Radio with Community	0.36 (0.48)	0.34 (0.47)	0.425
Ever Calls/Texts Radio Programs	0.49 (0.50)	0.49 (0.50)	0.911
News Consumption: Never (0-1)	0.28 (0.45)	0.29 (0.46)	0.720
News Consumption: Daily (0-1)	0.27 (0.44)	0.28 (0.45)	0.828
Watched Any TV Yesterday	0.23 (0.42)	0.26 (0.44)	0.352
Asset: Radio	0.43 (0.50)	0.46 (0.50)	0.404
Number of Radios	0.47 (0.58)	0.54 (0.67)	0.207
Asset: TV	0.14 (0.34)	0.19 (0.39)	0.074*
Asset: Cell Phone	0.85 (0.35)	0.79 (0.41)	0.114
Number of Rooms in Dwelling	3.22 (1.14)	3.27 (1.18)	0.530
Dwelling Has Mud Walls	0.75 (0.44)	0.76 (0.43)	0.831
Multiple Huts in Compound	0.11 (0.31)	0.13 (0.33)	0.394
Below-Average Living Conditions (Enum.)	0.40 (0.49)	0.31 (0.46)	0.352
Reject Forced Marriage	0.85 (0.36)	0.85 (0.36)	0.974
Gender Equality in Earning	0.42 (0.49)	0.46 (0.50)	0.172
Gender Equality in Raising Kids	0.90 (0.30)	0.87 (0.34)	0.494
Gender Equality in Leadership	0.85 (0.35)	0.89 (0.31)	0.163
No Preference for Sons	0.69 (0.46)	0.73 (0.44)	0.511
Openness to Change	0.69 (0.46)	0.70 (0.46)	0.954
Technology Is Good	0.97 (0.18)	0.96 (0.19)	0.808
Citizens Should Question Authority	0.30 (0.46)	0.34 (0.47)	0.344
Trust Elders' Judgement	0.80 (0.40)	0.78 (0.42)	0.485
Aware of HIV	1.00 (0.04)	1.00 (0.04)	0.974
Safe for HIV+ Person to Work with Kids	0.44 (0.50)	0.45 (0.50)	0.886
Comfortable on Bus with HIV+ Person	0.91 (0.28)	0.91 (0.29)	0.811
Knows Someone Excluded Socially (HIV)	0.06 (0.24)	0.05 (0.21)	0.386
Knows Someone Abandoned by Family (HIV)	0.05 (0.22)	0.04 (0.19)	0.336
Knows Someone Gossiped About (HIV)	0.56 (0.50)	0.63 (0.48)	0.132
Knows Someone Visited Less (HIV)	0.14 (0.34)	0.07 (0.26)	0.012**
Particularly Good Day Today	0.56 (0.50)	0.56 (0.50)	0.856

Note: Randomization p-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Standard deviations appear in parentheses next to the mean for each outcome. Full question text appears in the pre-registration; see Appendix subsection E.3.

Table A5: **Balance Table: Treatment and Placebo Among Screening Invitees, All Baseline Covariates**

Variable	Placebo (N=604)	Treatment (N=601)	RI p-value
Gender (1 = Female)	0.50 (0.50)	0.50 (0.50)	0.863
Age (years)	37.43 (12.07)	38.40 (11.92)	0.330
Married (1 = Married)	0.66 (0.47)	0.67 (0.47)	0.943
Religion: Muslim (1 = Muslim)	0.64 (0.48)	0.67 (0.47)	0.649
Religiosity (services/week)	5.04 (10.58)	4.66 (10.06)	0.721
Head of Household	0.58 (0.49)	0.57 (0.50)	0.669
Household Size	4.80 (2.38)	4.99 (2.22)	0.372
Household Members Aged 0–18	2.53 (1.67)	2.67 (1.77)	0.340
Children Ever Born	3.38 (2.79)	3.65 (2.86)	0.122
Education: Standard 7+	0.75 (0.43)	0.77 (0.42)	0.726
Literate	0.82 (0.38)	0.83 (0.38)	0.792
Primary Language: Swahili	0.81 (0.39)	0.80 (0.40)	0.735
Years Lived in Village	23.42 (15.05)	25.03 (14.64)	0.185
Lived in Village at Age 16	0.56 (0.50)	0.57 (0.50)	0.820
Village Knowledge (1–4)	2.70 (0.77)	2.68 (0.75)	0.827
Ever Lived in City/Town	0.46 (0.50)	0.46 (0.50)	0.876
Never Visits Big City	0.16 (0.37)	0.18 (0.38)	0.545
Never Visits Town	0.30 (0.46)	0.35 (0.48)	0.224
Radio Consumption (0–5)	1.33 (1.63)	1.25 (1.53)	0.321
Any Radio, Last 3 Months	0.64 (0.48)	0.66 (0.47)	0.438
Ever Listens to Radio with Community	0.37 (0.48)	0.35 (0.48)	0.645
Ever Calls/Texts Radio Programs	0.50 (0.50)	0.49 (0.50)	0.905
News Consumption: Never (0-1)	0.27 (0.45)	0.28 (0.45)	0.694
News Consumption: Daily (0-1)	0.27 (0.44)	0.29 (0.46)	0.512
Watched Any TV Yesterday	0.23 (0.42)	0.25 (0.43)	0.475
Asset: Radio	0.44 (0.50)	0.46 (0.50)	0.410
Number of Radios	0.49 (0.61)	0.55 (0.68)	0.208
Asset: TV	0.15 (0.35)	0.19 (0.39)	0.147
Asset: Cell Phone	0.85 (0.36)	0.80 (0.40)	0.193
Number of Rooms in Dwelling	3.20 (1.19)	3.28 (1.20)	0.400
Dwelling Has Mud Walls	0.75 (0.44)	0.75 (0.43)	0.899
Multiple Huts in Compound	0.11 (0.32)	0.13 (0.33)	0.557
Below-Average Living Conditions (Enum.)	0.37 (0.48)	0.31 (0.46)	0.585
Reject Forced Marriage	0.85 (0.35)	0.84 (0.36)	0.862
Gender Equality in Earning	0.41 (0.49)	0.45 (0.50)	0.164
Gender Equality in Raising Kids	0.90 (0.30)	0.87 (0.33)	0.513
Gender Equality in Leadership	0.85 (0.36)	0.87 (0.33)	0.403
No Preference for Sons	0.69 (0.46)	0.72 (0.45)	0.492
Openness to Change	0.68 (0.47)	0.69 (0.46)	0.716
Technology Is Good	0.95 (0.21)	0.96 (0.19)	0.576
Citizens Should Question Authority	0.32 (0.46)	0.33 (0.47)	0.639
Trust Elders' Judgement	0.80 (0.40)	0.78 (0.42)	0.452
Aware of HIV	1.00 (0.06)	1.00 (0.04)	0.933
Safe for HIV+ Person to Work with Kids	0.43 (0.50)	0.46 (0.50)	0.406
Comfortable on Bus with HIV+ Person	0.90 (0.30)	0.91 (0.29)	0.912
Knows Someone Excluded Socially (HIV)	0.07 (0.25)	0.05 (0.21)	0.252
Knows Someone Abandoned by Family (HIV)	0.05 (0.22)	0.04 (0.19)	0.419
Knows Someone Gossiped About (HIV)	0.57 (0.50)	0.62 (0.49)	0.202
Knows Someone Visited Less (HIV)	0.14 (0.35)	0.07 (0.26)	0.006***
Particularly Good Day Today	0.55 (0.50)	0.56 (0.50)	0.941

Note: Randomization p-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Standard deviations appear in parentheses next to the mean for each outcome. Full question text appears in the pre-registration; see Appendix subsection E.3.

Table A6: **Balance Table: Treatment and Placebo Among Endline Respondents, All Baseline Covariates**

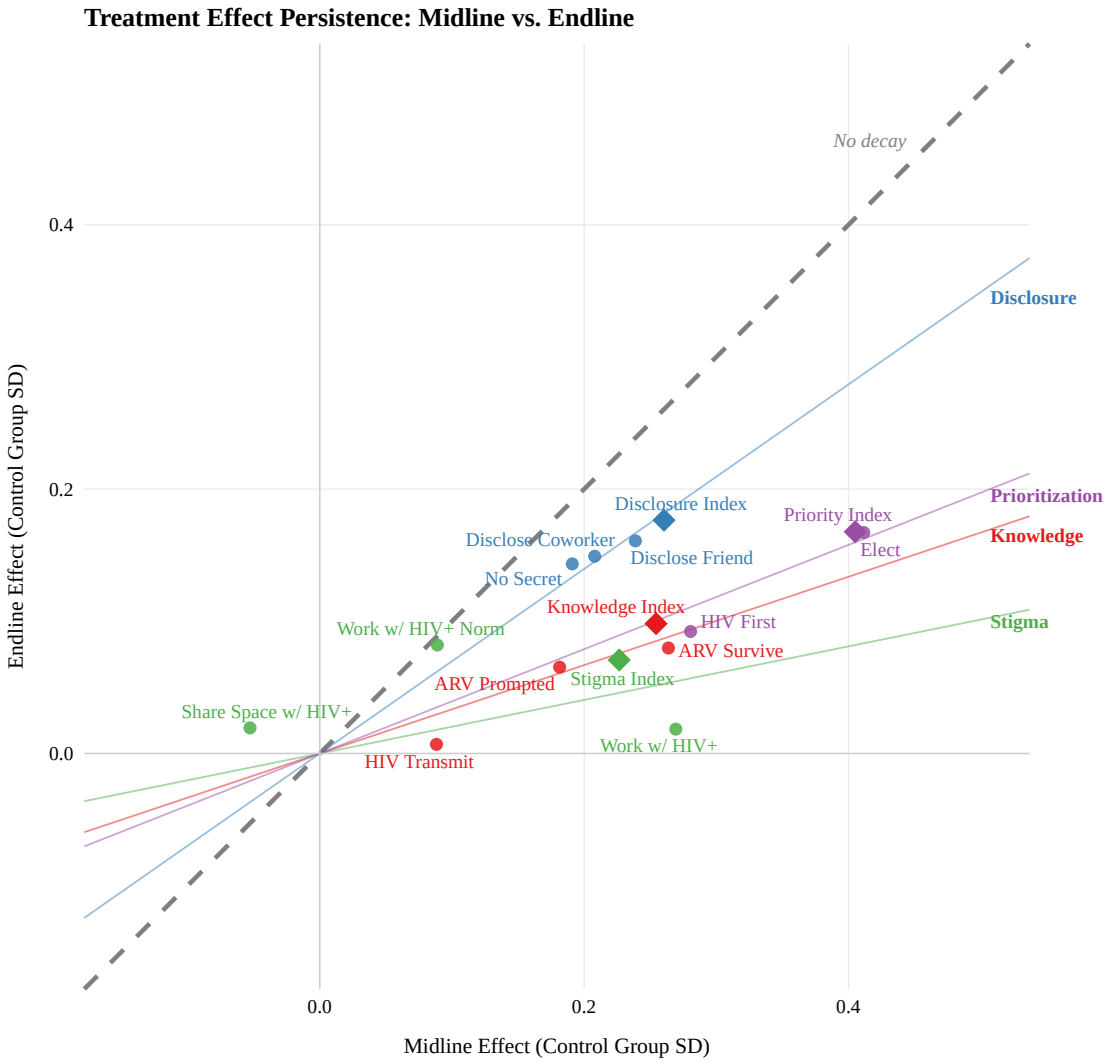
Variable	Placebo (N=582)	Treatment (N=581)	RI p-value
Gender (1 = Female)	0.50 (0.50)	0.50 (0.50)	0.848
Age (years)	37.51 (12.06)	38.35 (11.96)	0.386
Married (1 = Married)	0.66 (0.47)	0.67 (0.47)	0.875
Religion: Muslim (1 = Muslim)	0.64 (0.48)	0.67 (0.47)	0.596
Religiosity (services/week)	5.06 (10.61)	4.72 (10.13)	0.759
Head of Household	0.58 (0.49)	0.57 (0.50)	0.653
Household Size	4.82 (2.39)	5.02 (2.22)	0.404
Household Members Aged 0–18	2.53 (1.68)	2.69 (1.78)	0.330
Children Ever Born	3.40 (2.80)	3.64 (2.88)	0.196
Education: Standard 7+	0.76 (0.43)	0.77 (0.42)	0.759
Literate	0.83 (0.38)	0.83 (0.37)	0.841
Primary Language: Swahili	0.81 (0.39)	0.79 (0.41)	0.780
Years Lived in Village	23.51 (15.10)	25.24 (14.66)	0.176
Lived in Village at Age 16	0.56 (0.50)	0.57 (0.50)	0.838
Village Knowledge (1–4)	2.70 (0.77)	2.68 (0.75)	0.758
Ever Lived in City/Town	0.46 (0.50)	0.46 (0.50)	0.896
Never Visits Big City	0.15 (0.36)	0.18 (0.38)	0.398
Never Visits Town	0.30 (0.46)	0.35 (0.48)	0.189
Radio Consumption (0–5)	1.34 (1.64)	1.25 (1.54)	0.247
Any Radio, Last 3 Months	0.65 (0.48)	0.66 (0.47)	0.567
Ever Listens to Radio with Community	0.37 (0.48)	0.35 (0.48)	0.623
Ever Calls/Texts Radio Programs	0.49 (0.50)	0.50 (0.50)	0.984
News Consumption: Never (0-1)	0.28 (0.45)	0.28 (0.45)	0.820
News Consumption: Daily (0-1)	0.27 (0.44)	0.29 (0.46)	0.519
Watched Any TV Yesterday	0.24 (0.42)	0.25 (0.43)	0.537
Asset: Radio	0.43 (0.50)	0.46 (0.50)	0.358
Number of Radios	0.48 (0.59)	0.55 (0.68)	0.160
Asset: TV	0.14 (0.35)	0.19 (0.39)	0.058*
Asset: Cell Phone	0.86 (0.35)	0.80 (0.40)	0.168
Number of Rooms in Dwelling	3.18 (1.18)	3.29 (1.21)	0.278
Dwelling Has Mud Walls	0.75 (0.43)	0.75 (0.43)	0.928
Multiple Huts in Compound	0.12 (0.32)	0.13 (0.33)	0.643
Below-Average Living Conditions (Enum.)	0.37 (0.48)	0.31 (0.46)	0.574
Reject Forced Marriage	0.85 (0.36)	0.84 (0.36)	0.904
Gender Equality in Earning	0.41 (0.49)	0.45 (0.50)	0.125
Gender Equality in Raising Kids	0.90 (0.30)	0.87 (0.33)	0.564
Gender Equality in Leadership	0.85 (0.36)	0.87 (0.34)	0.427
No Preference for Sons	0.69 (0.46)	0.72 (0.45)	0.571
Openness to Change	0.68 (0.47)	0.69 (0.46)	0.738
Technology Is Good	0.96 (0.21)	0.96 (0.19)	0.568
Citizens Should Question Authority	0.31 (0.46)	0.33 (0.47)	0.620
Trust Elders' Judgement	0.80 (0.40)	0.77 (0.42)	0.456
Aware of HIV	1.00 (0.04)	1.00 (0.04)	0.968
Safe for HIV+ Person to Work with Kids	0.44 (0.50)	0.46 (0.50)	0.477
Comfortable on Bus with HIV+ Person	0.90 (0.30)	0.91 (0.29)	0.811
Knows Someone Excluded Socially (HIV)	0.07 (0.25)	0.04 (0.20)	0.175
Knows Someone Abandoned by Family (HIV)	0.05 (0.21)	0.03 (0.18)	0.388
Knows Someone Gossiped About (HIV)	0.56 (0.50)	0.62 (0.49)	0.143
Knows Someone Visited Less (HIV)	0.14 (0.35)	0.07 (0.26)	0.008***
Particularly Good Day Today	0.56 (0.50)	0.55 (0.50)	0.875

Note: * p < 0.10, ** p < 0.05, *** p < .01. Randomization p-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Standard deviations appear in parentheses next to the mean for each outcome. Full question text appears in the pre-registration; see Appendix subsection E.3.

D Supplementary Figures and Tables

D.1 Midline/Endline Comparison

Figure A1: **Midline vs Endline Treatment Effects: All Outcomes**



D.2 Spillover Appendix Components

D.2.1 Partner Spillover Appendix Components

Table A7: HIV Partner Knowledge Appendix

	Index	HIV Transmit	ARV No Spread	ARV Survive	ARV Prompted
Treatment Effect	0.009	0.034	0.019	0.006	-0.016
Standard Error	(0.01)	(0.017)	(0.01)	(0.019)	(0.019)
RI p-value	0.291	0.114	0.104	0.428	0.732
Test Direction	(+)	(+)	(+)	(+)	(+)
Placebo Mean	0.64	0.91	0.05	0.82	0.79
Placebo Village SD	0.06	0.08	0.04	0.11	0.1
DV Range	[1-0]	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00	0.00
Observations	709	709	708	708	705

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Missing covariate values were replaced with the mean value of the respondent's village. Column 1 reflects an index of the following columns. Column 2 was coded in response to the question, "To the best of your understanding, can a person with HIV/AIDS transmit the disease to the people that he or she works with?" If the respondent said that HIV could not be transmitted to the people that he or she works with, it was coded as a 1 and 0 if otherwise. Column 3 was coded in response to the question "In your understanding, what are the most important things a person with HIV/AIDS can do to make sure that they do not infect their spouse with HIV?" If the respondent mentioned ARVs, it was coded as a 1 and 0 if otherwise. Column 4 was coded in response to the question "In your understanding, what are the most important things a person with HIV/AIDS can do to help live longer?" If the respondent mentioned ARVs, it was coded as a 1 and 0 if otherwise. Column 5 was in response to the question, "Have you heard about antiretroviral drugs or ARVs that people infected with HIV can get from a doctor or a nurse to help them live longer?" If the respondent indicated that they heard about ARVs, it was coded as a 1 and 0 if otherwise.

Table A8: HIV Partner Disclosure Appendix

	Index	No Secret	Friend	Coworker
Treatment Effect	0.067	0.049	0.084	0.073
Standard Error	(0.026)	(0.034)	(0.024)	(0.026)
RI p-value	0.045	0.166	0.013	0.035
Test Direction	(+)	(+)	(+)	(+)
Placebo Mean	0.54	0.64	0.52	0.45
Placebo Village SD	0.14	0.15	0.14	0.17
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00
Observations	709	708	708	707

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Missing covariate values were replaced with the mean value of the respondent's village. Column 1 reflects an index of the following columns. Column 2 was in response to the question, "If a member of your family got HIV, would you want it to remain a secret?" If the respondent indicated that they would not want a family member's HIV diagnosis to remain a secret, it was coded as 1 and 0 if otherwise. Column 3 was coded in response to the question, "If you had HIV, would you tell or not tell your friends?" If the respondent said that they would tell their friends, it was coded as a 1 and 0 if otherwise. Column 4 was in response to the question "If you had HIV, would you tell or not tell your coworkers?" If the respondent said that they would tell their coworkers, it was coded as a 1 and 0 if otherwise.

Table A9: HIV Partner Stigma Appendix

	Index	Sit on Bus	Not Fired	Not Fired Norm
Treatment Effect	0	-0.003	-0.004	0.015
Standard Error	(0.014)	(0.021)	(0.013)	(0.018)
RI p-value	0.46	0.512	0.564	0.258
Test Direction	(+)	(+)	(+)	(+)
Placebo Mean	0.89	0.85	0.95	0.87
Placebo Village SD	0.05	0.1	0.05	0.08
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00
Observations	709	709	706	699

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Missing covariate values were replaced with the mean value of the respondent's village. Column 1 reflects an index of the following columns. Column 2 was coded in response to the question, "Would you be willing to sit next to someone with HIV/AIDS on the bus?" If the respondent would sit on the bus next to someone with HIV/AIDS, it was coded as a 1 and 0 if otherwise. Column 3 was in response to the question, "In your opinion, if a field worker on a large farm becomes infected with HIV, should they be fired or should they be allowed to keep their job?" If the respondent indicated that the worker should be fired, it was coded as a 0, and a 1 if the respondent indicated that the worker should be allowed to keep their job. Measuring what respondents believed their community's opinion was of Column 3, Column 4 was coded in response to the question "What do you think others in your community think about this issue?" If the respondent indicated that they believed others in the community believed the worker should be allowed to keep their job, it was coded as a 1 and 0 if otherwise.

Table A10: HIV Partner HIV Priority Appendix

	Index	HIV First	HIV Elect
Treatment Effect	0.05	0.095	0.006
Standard Error	(0.021)	(0.03)	(0.025)
RI p-value	0.072	0.024	0.444
Test Direction	(+)	(+)	(+)
Placebo Mean	0.32	0.1	0.53
Placebo Village SD	0.07	0.07	0.13
DV Range	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes
Covariates	No	No	No
Adjusted R2	0.00	0.00	0.00
Observations	709	709	709

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Missing covariate values were replaced with the mean value of the respondent's village. Column 1 reflects an index of the following columns. Column 2 reflects responses to the following request, "Here is another set of cards, which show different goals for your village. Please put them in order, from most important to you to least important to you." If respondents ranked "increasing the availability of medicine for HIV/AIDS" as their first priority, it was coded as a 1 and 0 if otherwise. Column 3 is in response to the following question, "The first candidate is named (name). (Name) is a (Muslim or Christian). (Name) promises villagers to make HIV/AIDS treatment more available in the village. The second candidate is named (name). (Name) is a (Muslim or Christian). (Name) promises villagers to (other platform was randomly selected). Which of these candidates do you think should be selected?" If respondents said that they would vote for the candidate who promises to expand HIV/AIDS treatment access, it was coded as 1 and 0 if otherwise.

D.2.2 Friend Spillover Appendix Components

Table A11: HIV Friend Knowledge and Disclosure Appendix

	Knowledge			Disclosure		
	Index	ARV Survive	ARV Prompted	Index	Friend	Coworker
Treatment Effect	0.004	0.006	0.002	0.01	0.028	-0.005
Standard Error	(0.029)	(0.033)	(0.031)	(0.022)	(0.026)	(0.021)
RI p-value	0.491	0.459	0.491	0.363	0.219	0.547
Test Direction	(+)	(+)	(+)	(+)	(+)	(+)
Placebo Mean	0.72	0.71	0.73	0.57	0.59	0.55
Placebo Village SD	0.13	0.19	0.12	0.14	0.14	0.15
DV Range	[1-0]	[1-0]	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00	0.00	0.00
Observations	972	957	972	972	972	970

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Missing covariate values were replaced with the mean value of the respondent's village. Column 1 reflects an index of the following columns. Column 2 was coded in response to the question "In your understanding, what are the most important things a person with HIV/AIDS can do to help live longer?" If the respondent mentioned ARVs, it was coded as a 1 and 0 if otherwise. Column 3 was in response to the question, "Have you heard about antiretroviral drugs or ARVs that people infected with HIV can get from a doctor or a nurse to help them live longer?" If the respondent indicated that they heard about ARVs, it was coded as a 1 and 0 if otherwise. Column 4 reflects an index of the following columns. Column 5 was coded in response to the question, "If you had HIV, would you tell or not tell your friends?" If the respondent said that they would tell their friends, it was coded as a 1 and 0 if otherwise. Column 6 was in response to the question "If you had HIV, would you tell or not tell your coworkers?" If the respondent said that they would tell their coworkers, it was coded as a 1 and 0 if otherwise.

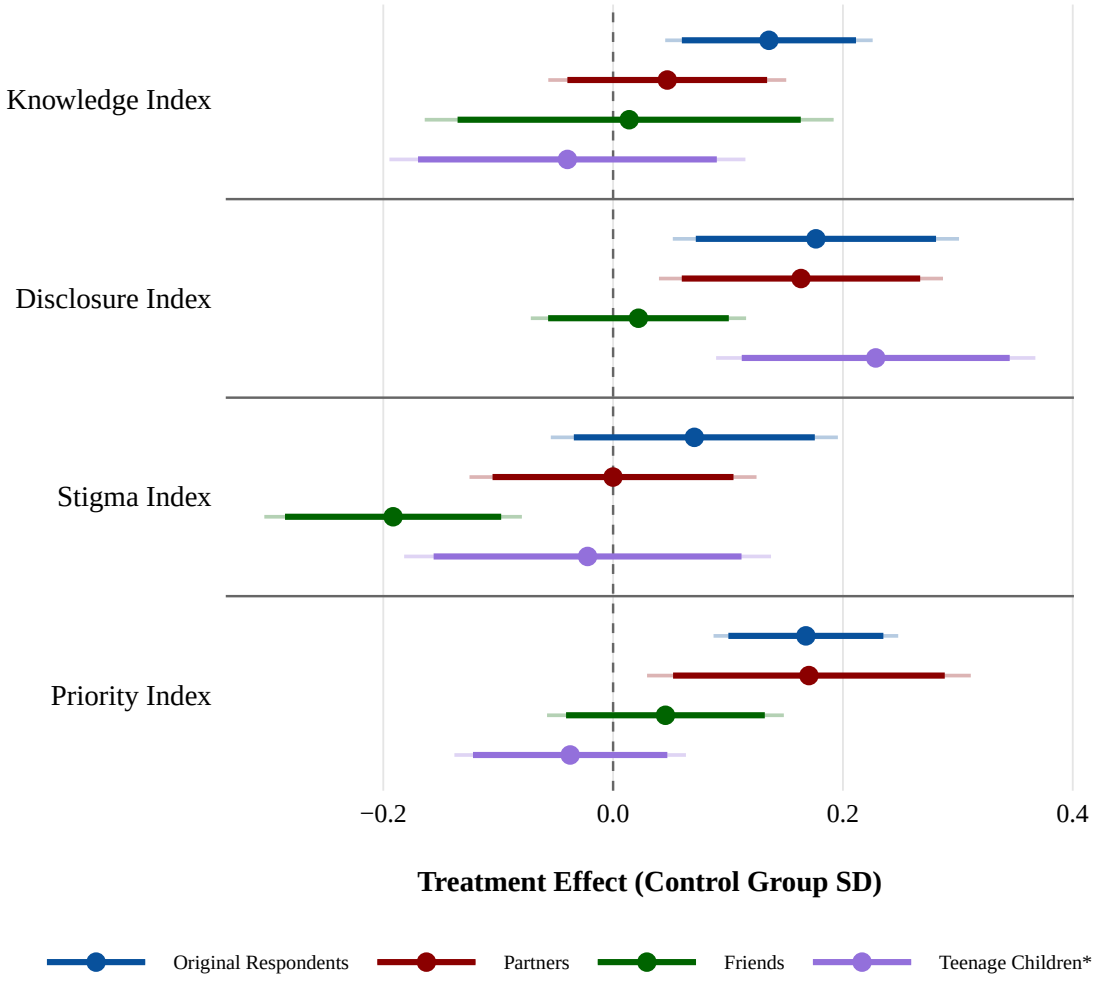
Table A12: HIV Friend Stigma and Priority Appendix

	Stigma	Priority		
	Sit on Bus	Index	HIV First	HIV Elect
Treatment Effect	-0.062	0.015	0.033	-0.005
Standard Error	(0.019)	(0.017)	(0.022)	(0.023)
RI p-value	0.979	0.291	0.194	0.583
Test Direction	(+)	(+)	(+)	(+)
Placebo Mean	0.88	0.36	0.15	0.57
Placebo Village SD	0.07	0.11	0.08	0.17
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00
Observations	972	972	972	971

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Missing covariate values were replaced with the mean value of the respondent's village. Column 1 was coded in response to the question, "Would you be willing to sit next to someone with HIV/AIDS on the bus?" If the respondent would sit on the bus next to someone with HIV/AIDS, it was coded as a 1 and 0 if otherwise. Column 2 reflects an index of the following columns. Column 3 reflects responses to the following request, "Here is another set of cards, which show different goals for your village. Please put them in order, from most important to you to least important to you." If respondents ranked "increasing the availability of medicine for HIV/AIDS" as their first priority, it was coded as a 1 and 0 if otherwise. Column 4 is in response to the following question, "The first candidate is named (name). (Name) is a (Muslim or Christian). (Name) promises villagers to make HIV/AIDS treatment more available in the village. The second candidate is named (name). (Name) is a (Muslim or Christian). (Name) promises villagers to (other platform was randomly selected). Which of these candidates do you think should be selected?" If respondents said that they would vote for the candidate who promises to expand HIV/AIDS treatment access, it was coded as 1 and 0 if otherwise.

D.3 Original Respondents versus Spillover Groups

Figure A2: **Original Respondents v Spillover Groups: Indices**



Note: *The HIV Stigma Index for Teenage Children included the following question that was not asked for original respondents, partners, or friends: "Would your parents advise you not to sit next to someone who has HIV on the bus? Or would they think it is ok?" If the respondent indicated that their parents would not advise them not to sit next to someone who has HIV on the bus, it was coded as 1 and 0 if otherwise.

Figure A3: **Original Respondents v Spillover Groups: Knowledge**

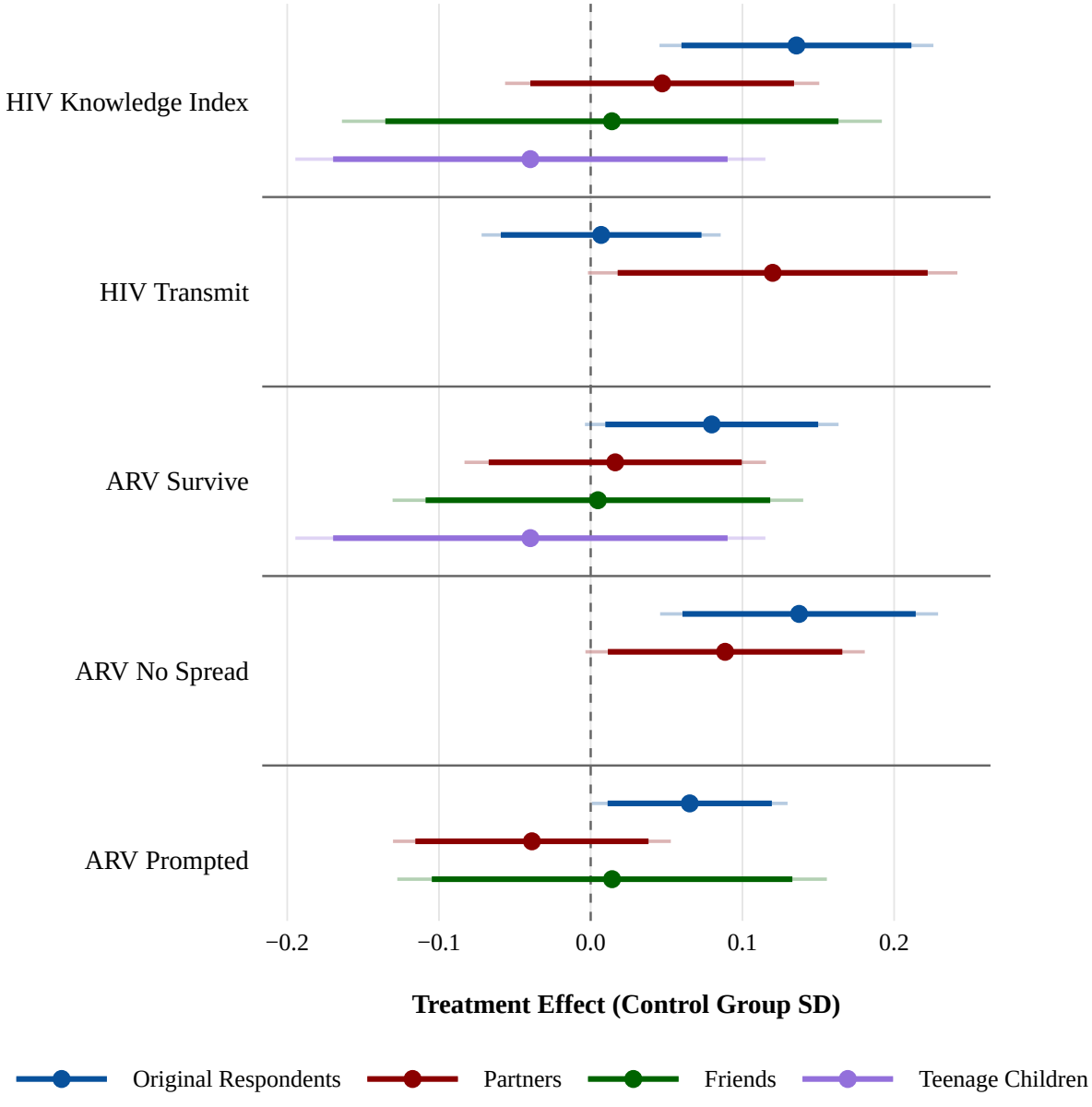


Figure A4: **Original Respondents v Spillover Groups: Disclosure**

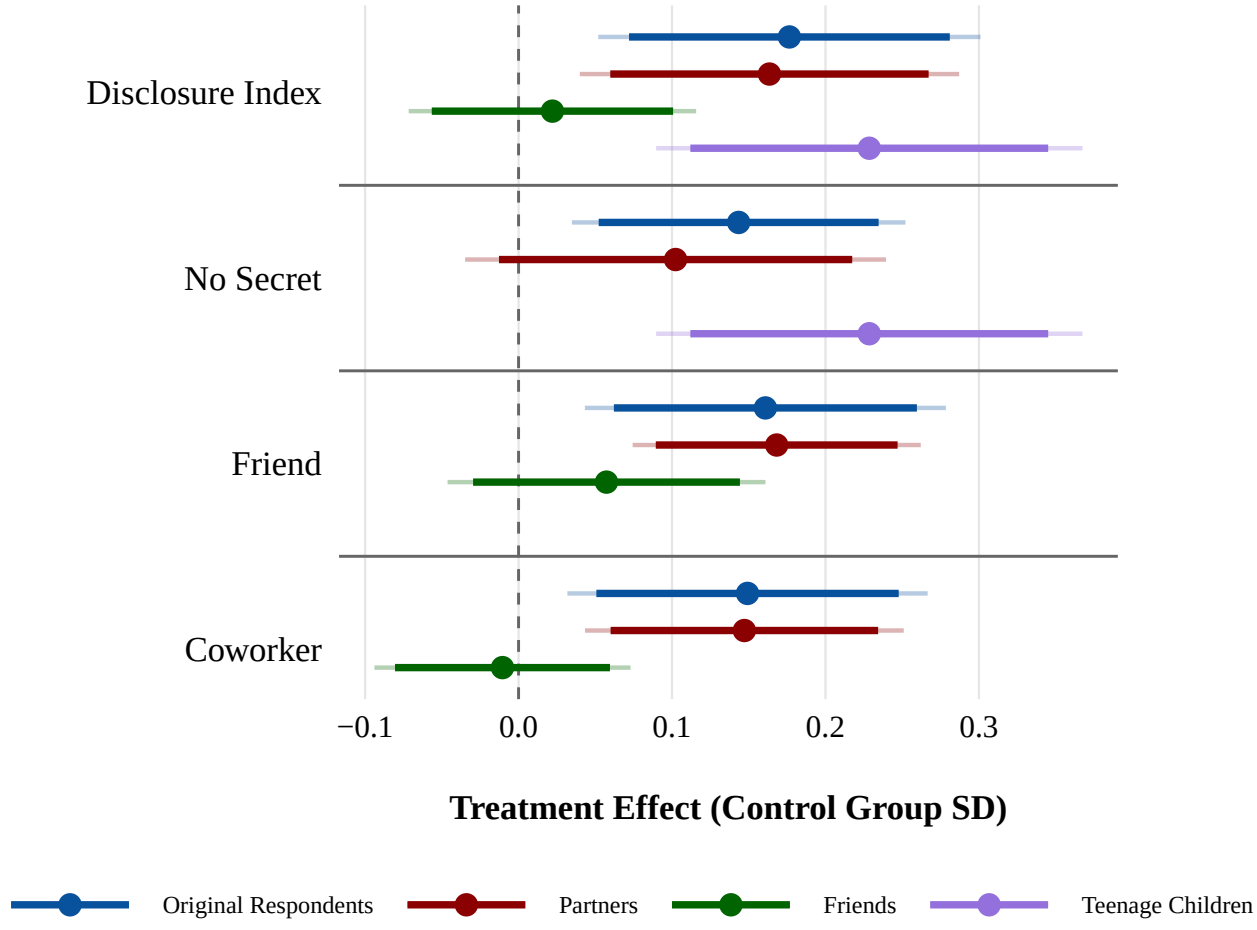


Figure A5: **Original Respondents v Spillover Groups: Stigma**

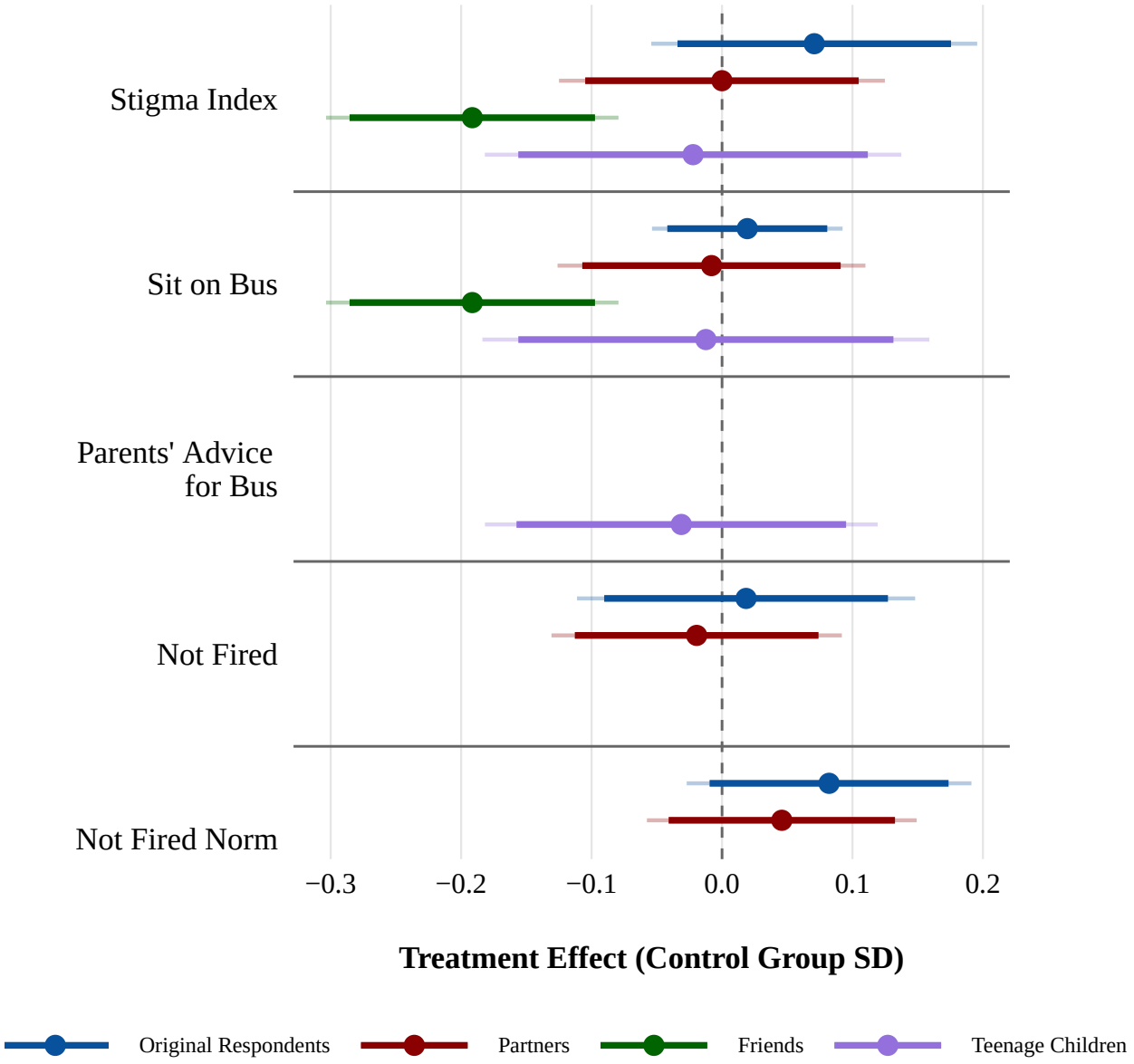
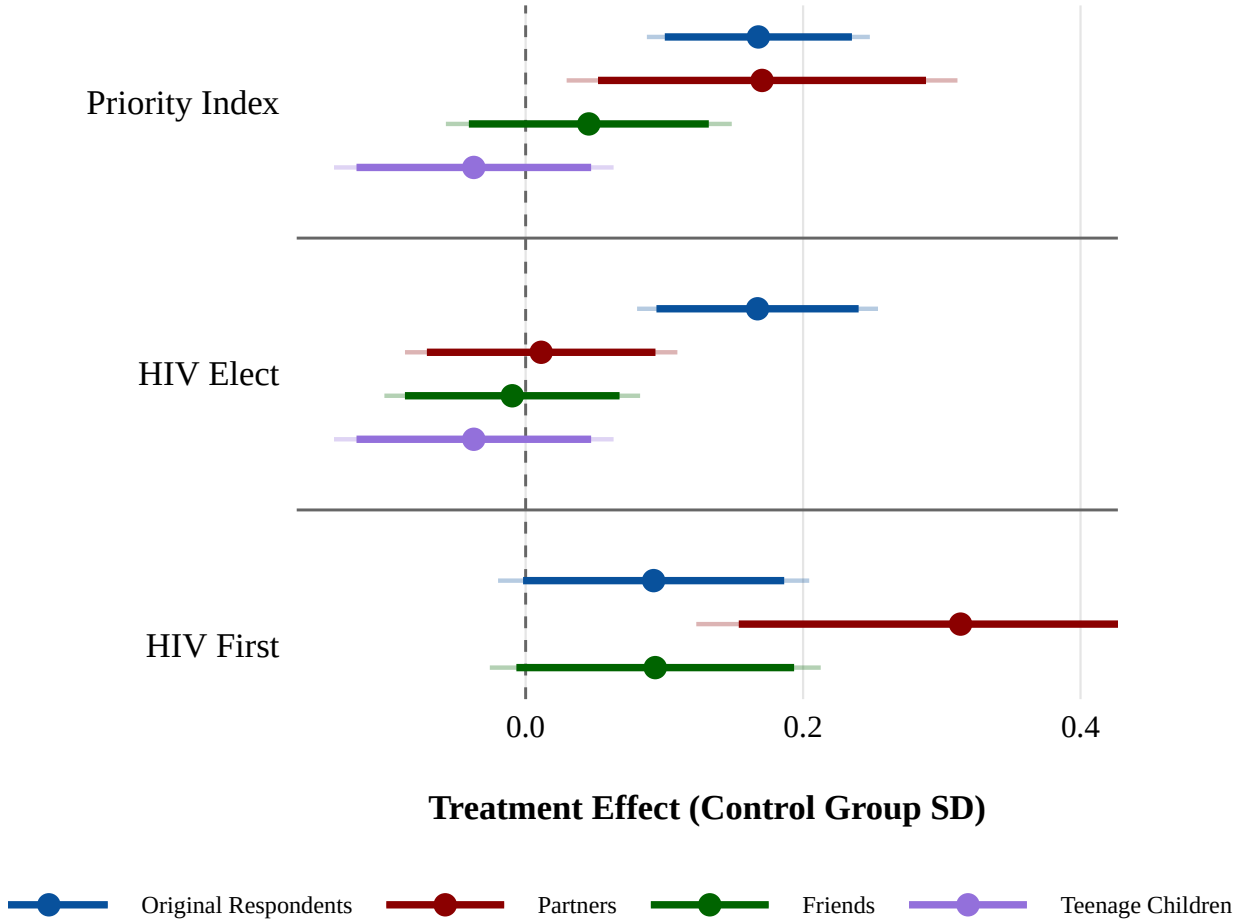


Figure A6: **Original Respondents v Spillover Groups: Prioritization**



D.4 Full Sample Results

D.4.1 Primary Respondents

Table A13: Full Sample: HIV Knowledge

	Index	Transmit	ARV Survive	ARV No Spread	ARV Prompted
Treatment Effect	0.019	0.008	0.034	0.015	0.003
Standard Error	(0.007)	(0.007)	(0.016)	(0.011)	(0.013)
RI p-value	0.044	0.228	0.078	0.167	0.433
Test Direction	(+)	(+)	(+)	(+)	(+)
Placebo Mean	0.61	0.95	0.82	0.05	0.84
Placebo Village SD	0.03	0.03	0.08	0.05	0.06
DV Range	[1-0]	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00	0.00
Observations	1,162	1,151	1,162	1,162	1,153

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Missing covariate values were replaced with the mean value of the respondent's village. Column 1 reflects an index of the following columns. Column 2 was coded in response to the question, "To the best of your understanding, can a person with HIV/AIDS transmit the disease to the people that he or she works with?" If the respondent said that HIV could not be transmitted to the people that he or she works with, it was coded as a 1 and 0 if otherwise. Column 3 was coded in response to the question "In your understanding, what are the most important things a person with HIV/AIDS can do to help live longer?" If the respondent mentioned ARVs, it was coded as a 1 and 0 if otherwise. Column 4 was coded in response to the question "In your understanding, what are the most important things a person with HIV/AIDS can do to make sure that they do not infect their spouse with HIV?" If the respondent mentioned ARVs, it was coded as a 1 and 0 if otherwise. Column 5 was in response to the question, "Have you heard about antiretroviral drugs or ARVs that people infected with HIV can get from a doctor or a nurse to help them live longer?" If the respondent indicated that they heard about ARVs, it was coded as a 1 and 0 if otherwise.

Table A14: Full Sample: HIV Disclosure

	Index	No Secret	Coworker	Friend
Treatment Effect	0.063	0.07	0.06	0.064
Standard Error	(0.024)	(0.025)	(0.028)	(0.027)
RI p-value	0.037	0.034	0.063	0.056
Test Direction	(+)	(+)	(+)	(+)
Placebo Mean	0.57	0.64	0.5	0.58
Placebo Village SD	0.17	0.15	0.2	0.18
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00
Observations	1,161	1,159	1,158	1,158

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Missing covariate values were replaced with the mean value of the respondent's village. Column 1 reflects an index of the following columns. Column 2 was in response to the question, "If a member of your family got HIV, would you want it to remain a secret?" If the respondent indicated that they would not want a family member's HIV diagnosis to remain a secret, it was coded as 1 and 0 if otherwise. Column 3 was in response to the question "If you had HIV, would you tell or not tell your coworkers?" If the respondent said that they would tell their coworkers, it was coded as a 1 and 0 if otherwise. Column 4 was coded in response to the question, "If you had HIV, would you tell or not tell your friends?" If the respondent said that they would tell their friends, it was coded as a 1 and 0 if otherwise.

Table A15: Full Sample: HIV Stigma

	Index	Sit on Bus	Not Fired	Not Fired Norm
Treatment Effect	0.008	0.005	0.01	0.025
Standard Error	(0.008)	(0.007)	(0.011)	(0.017)
RI p-value	0.227	0.315	0.237	0.127
Test Direction	(+)	(+)	(+)	(+)
Placebo Mean	0.96	0.97	0.96	0.89
Placebo Village SD	0.12	0.03	0.24	0.23
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00
Observations	1,162	1,159	1,159	1,155

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Missing covariate values were replaced with the mean value of the respondent's village. Column 1 reflects an index of the following columns. Column 2 was coded in response to the question, "Would you be willing to sit next to someone with HIV/AIDS on the bus?" If the respondent would sit on the bus next to someone with HIV/AIDS, it was coded as a 1 and 0 if otherwise. Column 3 was in response to the question, "In your opinion, if a field worker on a large farm becomes infected with HIV, should they be fired or should they be allowed to keep their job?" If the respondent indicated that the worker should be fired, it was coded as a 0, and a 1 if the respondent indicated that the worker should be allowed to keep their job. Measuring what respondents believed their community's opinion was of Column 3, Column 4 was coded in response to the question "What do you think others in your community think about this issue?" If the respondent indicated that they believed others in the community believed the worker should be allowed to keep their job, it was coded as a 1 and 0 if otherwise.

Table A16: Full Sample: HIV Priority

	Index	Elect	HIV First
Treatment Effect	0.048	0.069	0.032
Standard Error	(0.014)	(0.02)	(0.019)
RI p-value	0.019	0.021	0.172
Test Direction	(+)	(+)	(+)
Placebo Mean	0.33	0.51	0.13
Placebo Village SD	0.11	0.16	0.06
DV Range	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes
Covariates	No	No	No
Adjusted R2	0.00	0.00	0.00
Observations	1,163	1,163	1,123

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Missing covariate values were replaced with the mean value of the respondent's village. Column 1 reflects an index of the following columns. Column 2 is in response to the following question, "The first candidate is named (name). (Name) is a (Muslim or Christian). (Name) promises villagers to make HIV/AIDS treatment more available in the village. The second candidate is named (name). (Name) is a (Muslim or Christian). (Name) promises villagers to (other platform was randomly selected). Which of these candidates do you think should be selected?" If respondents said that they would vote for the candidate who promises to expand HIV/AIDS treatment access, it was coded as 1 and 0 if otherwise. Column 3 reflects responses to the following request, "Here is another set of cards, which show different goals for your village. Please put them in order, from most important to you to least important to you." If respondents ranked "increasing the availability of medicine for HIV/AIDS" as their first priority, it was coded as a 1 and 0 if otherwise.

D.4.2 Spillover Respondents

Table A17: Full Sample: Partner Spillover

	Knowledge	**Disclosure**	**Stigma**	**Priority**
	Index	Index	Index	Index
Treatment Effect	0.004	0.044	-0.017	0.024
Standard Error	(0.009)	(0.027)	(0.015)	(0.021)
RI p-value	0.365	0.134	0.76	0.221
Test Direction	(+)	(+)	(+)	(+)
Placebo Mean	0.65	0.54	0.9	0.32
Placebo Village SD	0.06	0.14	0.06	0.06
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00
Observations	816	816	816	816

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Missing covariate values were replaced with the mean value of the respondent's village. Column 1 reflects an index of the following HIV knowledge questions: HIV Transmit, ARV Survive, ARV No Spread, and ARV Prompted. The questions are the same as seen in Table A13. Column 2 reflects an index of the following HIV disclosure questions: No Secret, Disclosure to Coworker, and Disclosure to Friend. The questions are the same as seen in Table A14. Column 3 reflects an index of the following HIV stigma questions: Sit on Bus, Not Fired, and Not Fired Norm. The questions are the same as seen in Table A15. Column 4 reflects an index of the following HIV priority questions: HIV Elect and HIV First. The questions are the same as seen in Table A16.

Table A18: Full Sample: Friend Spillover

	Knowledge	**Disclosure**	**Stigma**	**Priority**
	Index	Index	Sit on Bus	Index
Treatment Effect	0.001	0.022	-0.053	0.012
Standard Error	(0.026)	(0.02)	(0.019)	(0.015)
RI p-value	0.504	0.198	0.971	0.308
Test Direction	(+)	(+)	(+)	(+)
Placebo Mean	0.73	0.56	0.88	0.36
Placebo Village SD	0.14	0.14	0.07	0.11
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00
Observations	1,109	1,109	1,109	1,109

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Missing covariate values were replaced with the mean value of the respondent's village. Column 1 reflects an index of the following HIV knowledge questions: ARV Survive and ARV Prompted. The questions are the same as seen in Table A13. Column 2 reflects an index of the following HIV disclosure questions: Disclosure to Coworker and Disclosure to Friend. The questions are the same as seen in Table A14. Column 3 was coded in response to the question, "Would you be willing to sit next to someone with HIV/AIDS on the bus?" If the respondent would sit on the bus next to someone with HIV/AIDS, it was coded as a 1 and 0 if otherwise. Column 4 reflects an index of the following HIV priority questions: HIV Elect and HIV First. The questions are the same as seen in Table A16.

Table A19: Full Sample: Teenage Children Spillover

	Knowledge	**Disclosure**	**Stigma**		**Priority**
	ARV Open-Ended	Family Keep Secret	Sit on Bus	Parents Advice for Bus	Elect
Treatment Effect	-0.009	0.074	-0.009	-0.006	-0.016
Standard Error	(0.031)	(0.042)	(0.04)	(0.04)	(0.04)
p-value	0.618	0.04	0.585	0.563	0.654
Test Direction	(+)	(+)	(+)	(+)	(+)
Placebo Mean	0.84	0.54	0.66	0.66	0.66
Placebo SD	0.36	0.5	0.47	0.47	0.47
DV Range	[1-0]	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	No	No	No	No	No
Covariates	No	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00	0.00
Observations	554	553	554	553	554

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Missing covariate values were replaced with the mean value of the respondent's village. Column 1 was coded in response to the question "In your understanding, what are the most important things a person with HIV/AIDS can do to help live longer?" If the respondent mentioned ARVs, it was coded as 1 and 0 if otherwise. Column 2 was coded in response to the question "If a member of your family got HIV, would you want it to remain a secret?" If the respondent indicated that they would not want it to remain a secret, it was coded as 1 and 0 if otherwise. Column 3 was coded in response to the question "Would you be willing to sit next to someone with HIV/AIDS on the bus?" If the respondent would sit on the bus next to someone with HIV/AIDS, it was coded as 1 and 0 if otherwise. Column 4 was coded in response to the question "Would your parents advise you not to sit next to someone who has HIV on the bus? Or would they think it is ok?" If the respondent indicated that their parents would not advise them not to sit next to someone who has HIV on the bus, it was coded as 1 and 0 if otherwise. Column 5 is in response to the following question, "The first candidate is named (name). (Name) is a (Muslim or Christian). (Name) promises villagers to make HIV/AIDS treatment more available in the village. The second candidate is named (name). (Name) is a (Muslim or Christian). (Name) promises villagers to (other platform was randomly selected). Which of these candidates do you think should be selected?" If respondents said that they would vote for the candidate who promises to expand HIV/AIDS treatment access, it was coded as 1 and 0 if otherwise.

E Supplementary Documents

E.1 Ethics

Our research protocols prioritized the welfare of participants. Our first concern was the psychological risks for HIV-positive individuals. We prioritized respect for participant autonomy throughout the study. Implementation teams provided detailed daily narratives documenting participant reactions and post-screening discussions to the principal investigators, which allowed the research team to monitor for any stigma-promoting effects or other unintended harms. Throughout the intervention period, these qualitative reports revealed no adverse participant responses to *Wahapahapa*. Our survey protocols also incorporated protective measures. We collaborated with Tanzanian research partners to craft culturally appropriate question wording. This collaborative process ensured that questions about HIV status and stigma avoided implicit pressure for status disclosure and minimized potential for emotional distress among respondents.

E.2 Reduced-Length Version of *Wahapahapa*

The *Wahapahapa* radio drama was written and developed by Media for Development International in collaboration with the Johns Hopkins Center for Communications Programs, with funding support from the Tanzanian Ministry of Health and Social Welfare, and the United States Agency for International Development (USAID) and the United States President's Emergency Fund for AIDS Relief (PEPFAR). *Wahapahapa* aired from 2007 to 2010.

The full version (approximately 120 hours) of *Wahapahapa* presented three interwoven storylines about a musical band in a fictional town in Tanzania. Each storyline communicates a different take-home message about HIV, and uses a transitional character to model targeted behaviors. *Wahapahapa's* storylines were tested extensively in focus group discussions before being included in the final plot.

The condensed version we present follows the story of Ray, an HIV-positive band member, and his decisions to seek ARV treatment and to disclose his HIV status to his family and coworkers.

Two additional storylines were not included in the condensed radio drama: a storyline emphasizing the value of parent-child communication around HIV and a storyline about the importance of confronting HIV stigma.

E.3 Pre-Analysis Plan

The pre-analysis plan registered on August 10, 2019 appears first, followed by the pre-analysis addendum covering the friend and child spillover samples, finalized before spillover data collection began.

Long-Term Audio Screening Pre-Analysis Plan

1. Pre-Analysis Plan

In this document we summarize key features and pre-analysis plan (PAP) for the second phase of the Community Media Randomized Trial in Tanga Region, Tanzania.

This document predates endline data collection, and our plan is therefore blind to outcomes. Any contingency not dealt with according to the Standard Operating Procedures for Don Green's lab at Columbia as of June 7, 2016.

2. Intervention Description

Two 1.5-2 hour radio dramas were produced and screened in rural Tanzania, where early and forced marriage and HIV/AIDS testing and treatment are prominent issues. The audio and translated scripts for the videos may be found in the digital materials that supplement the Community Radio PAP.

The treatment radio dramas are curated from two longer, multi-week radio soap operas produced by prominent non-governmental organizations based in Tanzania. *Tamapendo*, a radio drama produced by *Uzikwasa*, focuses on young community members challenging their community's traditional social norms about early forced marriage and gender-based violence. *Wahapahapa*, a radio drama produced by *Media for Development International*, focuses on characters challenging HIV/AIDS stigma and seeking appropriate treatment. The research team worked with the NGOs to shorten the radio dramas by focusing on one central plot line in each.

The *Tamapendo* plot line follows the journey of a young girl and her family as she tries to escape an early forced marriage arranged by her father. A supplementary plot line focuses on the effects of intimate partner violence on the girl's family. The *Wahapahapa* plot line follows the journey of a man coming to terms with his HIV positive status, opening up about his status to his family and coworkers, and seeking effective treatment.

3. Design and Randomization

3.1. Randomization

One key feature of the design is cluster randomization: audio screenings were held in 30 villages between July 25 and September 10, 2019. Villages were blocked into 15 wards and randomly assigned to either *Tamapendo* (focusing on early forced marriage) or *Wahapahapa* (focusing on HIV/AIDS stigma). The other is sampling within clusters: in each of 40 randomly selected households in each village, one randomly chosen household member (20 men and 20 women) was invited to attend the radio drama screening. Screenings were typically held in the early afternoon, and in some cases again in the evening. Participants were invited to the screenings at the end of the baseline interview. They were told that the screenings would feature free refreshments, and the screenings were free of charge. They were not told about the content of the program; enumerators were also blind to the village treatment condition.

3.2. Site Selection and Blocking

We conducted an initial census of wards (as defined by the 2018 Tanzanian Census) with at least two villages that met the following criteria:

- At least 70 km from Pangani town[1]
- No main or secondary road running through the village polygon
- At least 8 km from any town or city
- Neither in Pangani District, nor in our Natural Experiment sample
- At least 100 households within an 800 meter radius from the village center, as estimated from satellite images
- At most 1,000 households within a 400 meter radius, as estimated from satellite imagines
- At least two villages 4 km from one another in the ward, as measured in QGIS

[1] This condition was important for a radio-distribution experiment that we conducted in parallel to the radio screening experiment, because radio recipients needed to be within range of Pangani FM.

We identified 16 wards with 45 villages meeting the target criteria. We randomly selected two villages in wards containing three or more eligible villages and preserved the remaining wards in a random order for selection in the event that a village needed to be discarded during data collection. The code for sampling villages is included in Appendix 1 of the Community Radio PAP.

Pangani FM – Population Selection

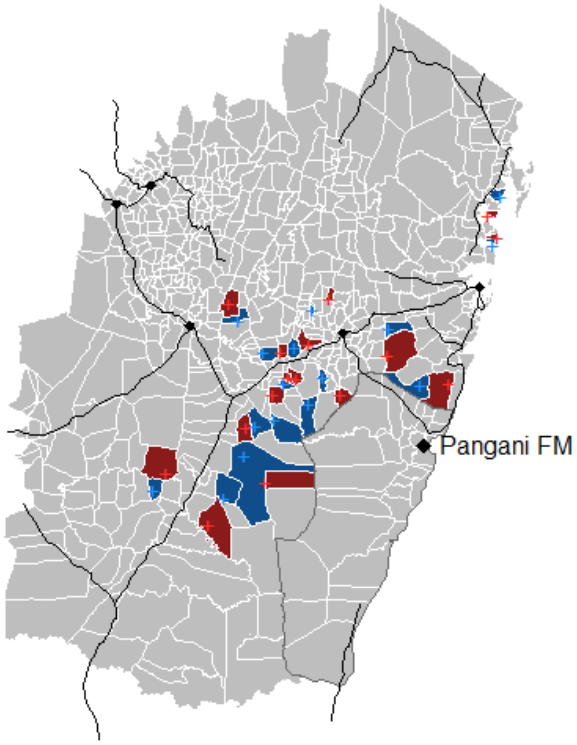
3.3. Random assignment

The sampling process described in section 3.2 yielded 16 blocks (15 blocks for treatment / control and one for replacement) of 2 clusters each.

We then randomly assigned Treatment and Placebo screening conditions within each pair. In the case that a village needed to be discarded, we specified that we would replace the village with the next village on the ward's list of randomly ordered. We conducted randomization after confirming two viable villages in a ward. In the case that the ward had no new villages meeting the criteria specified in Section 3.2, we moved to the replacement ward.

The randomization code is submitted with this Pre-Analysis Plan.

Pangani FM – Random Selection of Villages



Blue = Treatment, Red = Control, + = Village center

3.4. Sampling and subjects

In the baseline survey, we conducted 1,205interviews with respondents between 18 and 65 years of age (601 male and 605 female respondents).

Within each village, households are sampled from a circular area around the village center. The village center was identified using satellite images in QGIS. The research team picked an appropriate radius between 400, 600 and 800 meters for each village based on the population density of the village inferred from satellite images.

The mobilization team received a high-resolution satellite image of the village with the village center and appropriate radius pre-identified on Avenza Maps. They then worked with village leaders (usually the Village Executive Officer or Village Chairperson) to produce input a list of all households that have resided within the designated radius for at least six months into SurveyCTO. SurveyCTO then generated a list of 40 randomly selected households from the village census (20 randomly selected for female respondents and 20 randomly selected for male respondents) as well as 20 substitute households.

1-2 days after the mobilization team conducted the household listing, the survey team visited the 40 targeted households. Surveyors and respondents were matched by gender. Upon reaching the household, the enumerator listed all household members of the designated gender between the ages of 18 and 65, and randomly selected one of them as the respondent.

In cases in which no household member was of the designated gender and age category, or in which the respondent was unable or unwilling to complete the survey, the survey team replaced them with the next randomly selected household from the original village-level randomization form.

Only the 40 villagers who completed the survey were invited to the village screening. All 40 respondents were targeted for the first follow-up survey. The midline pre-analysis plan specified that if more than four respondents were unavailable during the follow-up survey in any village half of the unavailable respondents would be randomly selected for intensive measurement efforts and used to represent the full set of missing respondents in that village. However, the team did not reach this level of attrition in any village, so the approach was not used.

We will target the same list of 40 respondents in each village for the endline survey, and apply the same rule for dealing with attrition: if more than 4 respondents are missing from any village, we will randomly select half of the unavailable respondents for intensive tracking and used to represent the full set of missing respondents in that village.

4. Analysis of Spillovers

We decided to explore not only the direct effects of the audio screenings on respondents' outcomes, but also the potential spillover effects that arise through respondents' discussion with their partners, children, and community. We will conduct the partner spillover survey in parallel to the primary endline survey.

This PAP describes the partner spillover survey, which will be conducted in parallel to the primary endline survey. Future PAPs will discuss the sampling, survey, and analysis strategy for the community and child spillover survyes, which will be conducted in January 2021.

4.1. Partner Spillover Survey

4.1.1. Sampling

We conduct a partner survey with one partner of each subject who reported being married or "living as married" in the 2019 baseline survey. The mobilization team will approach each respondent and ask for the name and location of the partner that they spend the most time with. Where a respondent reports spending equal time with more than one spouse, we will take information about the spouse who the partner met "first".

4.1.2. Expectations

While we will field separate surveys to measure spillover effects, we include a number of relevant questions in this survey to facilitate future data collection and analysis. First, we ask respondents about relationship roles and division of responsibilities in the household, as well as topics they tend to discuss with their spouse, including child marriage and HIV/AIDS; relationship style and the frequency and depth of interactions and conversations between partners could affect the transmission of information from the audio screening from a respondent to his or her spouse. We expect relationship style and the frequency and depth of interactions to moderate spillovers to partners; large effects are expected among those with whom the respondent engages in discussion about relevant topics. We expect spillovers to be primarily concentrated among partners of individuals who attended the audio screening.

5. Primary Hypotheses

Our outcomes for the long-term radio screening experiment are drawn from the endline survey conducted one year and three months after the village screening.

The outcomes of interest are listed here for the relevant samples and with reference to the two shows, Tamapendo (T) and Wahapahapa (W).

	Endline	**Partner Spillover**
Compliance	**yes**	**yes**
Knowledge: • W: Increased knowledge of HIV treatment and transmission	**yes**	**yes**
Attitudes • T: Reduced support for early and forced marriage • W: Reduced stigma against HIV positive individuals • W: Increased support for HIV disclosure	**yes**	**yes**
Perceived norms • T: Increased perception of community rejection of early marriage; increased perception of community's willingness to report EM; increased perception of partner rejection of forced marriage and prioritization of ending EM • W: Decreased perception of HIV stigma in community	**yes**	**yes**
Behavioral Intentions • T: Increased willingness to report instances of EM to authorities; increased willingness for women to resist EM • W: Increased willingness to reveal HIV/AIDS status to friends/family	**yes**	**yes**
Behavioral outcomes • T: Increased willingness to speak out (and record a message) about early marriage, increased willingness to share this message with local leaders	**yes**	**yes**
Political priorities	**yes**	**yes**

T*:* Votes for hypothetical candidate committed to reducing early forced marriage; increased support for women's political participation; increased priority ranking of "early forced marriage" among nine policy objectives for the village • W*:* Votes for hypothetical candidate committed to improving access to HIVE/AIDS treatment; increased priority ranking of "early forced marriage" among nine policy objectives for the village		
Gender Equality • T: Increased support for gender equality • T: Increased rejection of intimate partner violence ▪ T: Increased perception of community rejection of intimate partner violence ▪ T: Increased willingness to report intimate partner violence	**yes**	**yes**

Note that for some of the outcomes (rejection of EM, reporting EM, discrimination based on HIV) we also collect each partner's perception of the other partner's stance on the matter.

5.1. Survey Experiments

Within this data collection wave, we embedded a number of survey experiments.

4.2.1. Elite Cues (in both endline and partner spillover surveys)

The first survey experiment seeks to identify the effect of elite position-taking on respondents' support for women's political participation (WPP). The pre-analysis plan for this survey experiment, including hypotheses, is contained in the Pangani FM Natural Experiment pre-analysis plan.

4.2.2. Pluralistic Ignorance (only in endline survey)

The second survey experiment seeks to understand the role of the mismatch between first- and second- order beliefs of the respondents around Early Marriage. The pre-analysis plan for this survey experiment, including hypotheses, is contained in the Pangani FM Natural Experiment pre-analysis plan.

4.2.3. Laws and Persuasion (only in the partner spillover survey)

This survey experiment aims to explore the effects of knowledge about laws regarding early marriage (EM) on respondents' attitudes and behaviors. It is unclear whether revisions to EM laws cause citizens to change their attitudes and behaviors, or whether citizens remain anchored to their prior beliefs, community traditions, and elite opinions even after learning about EM laws. To determine whether citizens respond to information about formal laws promulgated by Tanzania's legal institutions, we randomly assign respondents to one of three treatments:

1. Treatment 1: One-sided information about the Tanzanian Court of Appeals' recent anti-EM ruling. Respondents are informed that the Court's ruled in 2019 that it was a legal offense for a girl to marry under the age of 18, even if the girl's religion or culture allow it.
2. Treatment 2: Two-sided information about the Tanzanian Court of Appeals' recent anti-EM ruling. Respondents are informed that (1) the Tanzanian High Court prohibited marriage under the age of 18 in 2016, (2) the Attorney General challenged the ruling on the grounds that there should be a religious exemption, and (3) the Court of Appeals upheld the ruling that marriage under the age of 18 was illegal, even if the girl's religion or culture allow it.
3. Control: No information is shared.[2]

The outcomes of interest focus on the acceptability of Early Marriage given a range of hypothesized variations of the situation (the religion of the girl; the girl's likelihood of getting pregnant; the girl's educational and future marriage prospects; and the extent of economic hardship experienced by the girl's family). We are also interested in the effect of the treatment on perceive acceptability of early marriage in the community. Also varied was the cost to the respondent (the level of observability) of taking action to signal their rejection of early marriage. This behavioral measure consists of asking respondents whether they are willing to record an audio message that expresses their position on Early Marriage, either for or against, and then whether they would like to add their name to it, and whether they would like to share it with the District Council of the District and/or Pangani FM, a local radio station.

We will examine the treatment effect of each of the two treatments independently through a simple differences-in-means (OLS) as well as a covariate-adjusted regression, using lasso regression to select the minimal number of covariates that best predict each outcome. We predict that we will positive effects on anti-EM attitudes and behaviors for the one-sided message treatment that are attenuated by the

[2] Respondents in the control group are informed of the decision later in the survey

two-message treatment; however, it is unclear whether the two-message treatment leads to greater or lesser support compared to the control condition. Moreover, we are agnostic about whether the effect will be the same across attitudinal and behavioral outcomes.

6. Estimation

6.1. General Coding Rules

- Unless otherwise indicated, outcome measures will be constructed from single items
- We will construct variables so that higher values represent more agreement with the norms portrayed in the radio screenings.
- Where more than one survey response relates to a common outcome construct, we will generate an index in the following manner: all variables will be standardized to range from 0 to 1 by dividing by the maximum possible value. All variables will then be summed and divided by the number of variables in the index.

6.2. Compliance

We expect that the effects of screening will be larger for baseline respondents who attended the screening, as compared to noncompliers who did not attend. The placebo design allows us to compare just compliers in both experimental groups. The midline results confirmed that compliance is not significantly related to assigned treatment; we will check to make sure that this finding persists in the endline given that some attrition will occur in our survey panel. We will also check balance to verify that compliers' partners have similar profiles in both experimental groups.

6.3. Primary Specification

For all analyses, we will estimate treatment effects using least squares regression and clustered standard errors, with fixed effects for block (ward). Additional covariates will be selected using the process outlined in Section 5.6.

For example, restricting the sample to compliers' partners, we would assess the (spillover) effects of the treatment on partners' attitudes about forced marriage by regressing this outcome on a treatment indicator and block indicators.

6.4. Basic Specification

For all analyses, we will also estimate treatment effects using only fixed effects for block (ward) and no other covariates.

6.5. Randomization Inference

For all specifications, we will calculate p-values with randomization inference using the randomization code provided in Appendix 2. We will calculate cluster-robust standard errors using the sandwich package for R but will not rely on these standard errors for inference.

6.6. Covariates

As noted above, we will include an indicator for block (ward) in all specifications.

We will use lasso to select the minimal number of covariates that best predict each outcome and include only these covariates in our estimation. In our "agnostic" specification, we will draw potential covariates exclusively from the baseline survey. In our "extensive" specification, we include covariates measured in the endline that, although posttreatment, are not plausibly affected by our interventions. These include measures of political interest, political participation, and psychological traits identified in the literature as promoting narrative persuasion (openness, need for cognition, and absorption).

We will perform lasso on covariate data that has been pre-processed to remove item-level missingness through multiple chain equations, as described below. The lasso procedure that we plan to use features a generalized linear model with lasso penalization and is implemented in the glmnet package for R. The loss function requires selecting a regularization parameter that determines the severity of the penalty for including extra covariates. Since this regularization parameter cannot be optimally chosen in advance, we will select it using 10-fold cross-validation. Specifically, for each outcome, we will choose the regularization parameter that minimizes the 10-fold cross-validation error averaged over 10 runs (since the folds are chosen at random). Only the covariates retained by the lasso will be included in the specification.

We will perform this lasso variable selection method using the entire list of baseline covariates, including block and resample indicators, among the entire sample. For extensive specification, the process will be repeated for the expanded set of covariates. Note that this procedure implies that for each outcome, a different number of covariates may be included on the right-hand side.

6.7. Heterogeneous Treatment Effects

We will explore individual-level treatment heterogeneity using the generalized causal forest algorithm in R, as we have done for the midline results. When the sources of treatment effect heterogeneity are not known *ex ante*, restricting analyses to pre-registered hypotheses may hinder the discovery of unanticipated results; however, ad hoc exploration of treatment effect heterogeneity runs the risk of false discovery. We therefore use machine learning tools originally developed to predict individualized responses to medical interventions in order to automate and synthesize the search for treatment effect heterogeneity across many covariates.

For the purpose of this analysis we will use the large set of covariates collected at baseline, as well as other post-treatment covariates that are plausibly unaffected by treatment status. These plausibly exogenous attributes include parenting style, partner relationship measures (especially discussion), levels of political interest and participation, and psychological traits.

6.8. Multiple Comparisons

We address the problem of multiple comparisons in the following way. Estimated t-ratios for the analyses will be sorted in descending order for all one-tailed hypothesis. A null distribution of t-ratios will be simulated by repeating the random assignment 10,000 times and for each simulation generating a vector of estimated t-ratios. To obtain the p-value for the largest estimated t-ratio, we will ask: what proportion of the simulations generated a largest p-value at least as large as what we obtained empirically? The same procedure will be applied to the second largest p-value, and so forth (Anderson 2008).

6.9. Item Level Missingness

Non-response to outcome questions will be dealt with through multivariate imputation via chained equations (MICE) using the mice package for R. Imputations for outcome measures will be performed within issue areas (IPV, HIV, EM).

Addendum to Long-Term Audio Screening Pre-Analysis Plan: Community Spillover and Children Spillover

1. Pre-Analysis Plan

In this document we summarize key features and pre-analysis plan (PAP) for the final phase of the Community Media Randomized Trial in Tanga Region, Tanzania.

In December 2020, we completed endline data collection for the original respondents in the Audio Screening experiment as well as their partners. This document pertains to the samples consisting of the children and close friends of original respondents; we seek to measure spillover effects from the original respondents to their children and close friends, This document predates data collection for the "spillover" portion of the survey and our plan is therefore blind to outcomes. Any contingency not dealt with according to the Standard Operating Procedures for Don Green's lab at Columbia as of June 7, 2016.

2. Intervention Description

Two 1.5-2 hour radio dramas were produced and screened in rural Tanzania, where early and forced marriage and HIV/AIDS testing and treatment are prominent issues. The audio and translated scripts for the videos may be found in the digital materials that supplement the Community Radio PAP.

The treatment radio dramas are curated from two longer, multi-week radio soap operas produced by prominent non-governmental organizations based in Tanzania. *Tamapendo*, a radio drama produced by *Uzikwasa*, focuses on young community members challenging their community's traditional social norms about early forced marriage and gender-based violence. *Wahapahapa*, a radio drama produced by *Media for Development International*, focuses on characters challenging HIV/AIDS stigma and seeking appropriate treatment. The research team worked with the NGOs to shorten the radio dramas by focusing on one central plot line in each.

The *Tamapendo* plot line follows the journey of a young girl and her family as she tries to escape an early forced marriage arranged by her father. A supplementary plot line focuses on the effects of intimate partner violence on the girl's family. The *Wahapahapa* plot line follows the journey of a man coming to terms with his HIV positive status, opening up about his status to his family and coworkers, and seeking effective treatment.

3. Design and Randomization

3.1. Randomization

One key feature of the design is cluster randomization: audio screenings were held in 30 villages between July 25 and September 10, 2019. Villages were blocked into 15 wards and randomly assigned to either *Tamapendo* (focusing on early forced marriage) or *Wahapahapa* (focusing on HIV/AIDS stigma). The other is sampling within clusters: in each of 40 randomly selected households in each village, one randomly chosen household member (20 men and 20 women) was invited to attend the radio drama screening. Screenings were typically held in the early afternoon, and in some cases again in the evening. Participants were invited to the screenings at the end of the baseline interview. They were told that the screenings would feature free refreshments, and the screenings were free of charge. They were not told about the content of the program; enumerators were also blind to the village treatment condition.

3.2. Site Selection and Blocking

We conducted an initial census of wards (as defined by the 2018 Tanzanian Census) with at least two villages that met the following criteria:

- At least 70 km from Pangani town[1]
- No main or secondary road running through the village polygon
- At least 8 km from any town or city
- Neither in Pangani District, nor in our Natural Experiment sample

[1] This condition was important for a radio-distribution experiment that we conducted in parallel to the radio screening experiment, because radio recipients needed to be within range of Pangani FM.

- At least 100 households within an 800 meter radius from the village center, as estimated from satellite images
- At most 1,000 households within a 400 meter radius, as estimated from satellite imagines
- At least two villages 4 km from one another in the ward, as measured in QGIS

We identified 16 wards with 45 villages meeting the target criteria. We randomly selected two villages in wards containing three or more eligible villages and preserved the remaining wards in a random order for selection in the event that a village needed to be discarded during data collection. The code for sampling villages is included in Appendix 1 of the Community Radio PAP.

Pangani FM – Population Selection

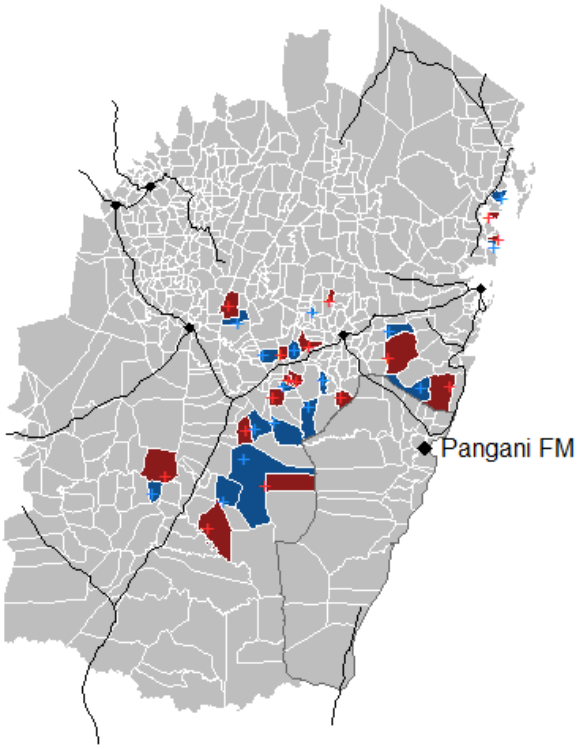
3.3. Random assignment

The sampling process described in section 3.2 yielded 16 blocks (15 blocks for treatment / control and one for replacement) of 2 clusters each.

We then randomly assigned Treatment and Placebo screening conditions within each pair. In the case that a village needed to be discarded, we specified that we would replace the village with the next village on the ward's list of randomly ordered. We conducted randomization after confirming two viable villages in a ward. In the case that the ward had no new villages meeting the criteria specified in Section 3.2, we moved to the replacement ward.

The randomization code is submitted with this Pre-Analysis Plan.

Pangani FM – Random Selection of Villages



Blue = Treatment, Red = Control, + = Village center

3.4. Sampling and subjects

In the baseline survey, we conducted 1,205 interviews with respondents between 18 and 65 years of age (601 male and 605 female respondents).

Within each village, households are sampled from a circular area around the village center. The village center was identified using satellite images in QGIS. The research team picked an appropriate radius between 400, 600 and 800 meters for each village based on the population density of the village inferred from satellite images.

The mobilization team received a high-resolution satellite image of the village with the village center and appropriate radius pre-identified on Avenza Maps. They then worked with village leaders (usually the Village Executive Officer or Village Chairperson) to produce input a list of all households that have resided within the designated radius for at least six months into SurveyCTO. SurveyCTO then generated a list of 40 randomly selected households from the village census (20 randomly selected for female respondents and 20 randomly selected for male respondents) as well as 20 substitute households.

1-2 days after the mobilization team conducted the household listing, the survey team visited the 40 targeted households. Surveyors and respondents were matched by gender. Upon reaching the household, the enumerator listed all household members of the designated gender between the ages of 18 and 65, and randomly selected one of them as the respondent.

In cases in which no household member was of the designated gender and age category, or in which the respondent was unable or unwilling to complete the survey, the survey team replaced them with the next randomly selected household from the original village-level randomization form.

Only the 40 villagers who completed the survey were invited to the village screening. All 40 respondents were targeted for the first follow-up survey. The midline pre-analysis plan specified that if more than four respondents were unavailable during the follow-up survey in any village half of the unavailable respondents would be randomly selected for intensive measurement efforts and used to represent the full set of missing respondents in that village. However, the team did not reach this level of attrition in any village, so the approach was not used.

We will target the same list of 40 respondents in each village for the endline survey, and apply the same rule for dealing with attrition: if more than 4 respondents are missing from any village, we will randomly select half of the unavailable respondents for intensive tracking and used to represent the full set of missing respondents in that village.

4. Analysis of Spillovers

We decided to explore not only the direct effects of the audio screenings on respondents' outcomes, but also the potential spillover effects that arise through respondents' discussion with their partners, their children, and the community. The partner spillover survey was conducted in December 2020 in parallel with the primary endline survey.

The primary Long-Term Audio Screening Pre-Analysis Plan describes the partner spillover survey, which was conducted in parallel with the primary endline survey. This document is an addendum to the original PAP. It discusses the sampling, survey, and analysis strategy for the community and child spillover surveys, which will be conducted in January-February 2021.

4.1. Community Spillover Survey

In January-February 2021, we will field a separate survey to measure spillover effects to community members outside of respondent households.

4.1.1. Sampling

During the primary endline survey in December 2020, enumerators asked respondents to list individuals outside of respondents' household that the respondent felt closest to. We excluded subjects who:

1. Were close family members of the respondent (parent, child, grandparent, or aunt/uncle)
2. Were not between the ages of 18 and 65
3. Did not live in the village of the respondent
4. Had attended the audio screening

The mobilization team will approach each respondent and ask for the location of the listed friends (and up to two additional friends if the original friends were excluded for the reasons listed above). Starting with the closest friend, the mobilizer will visit the friend and ask them to participate in a survey about community issues. If the "best" friend is not available, the mobilizer will move to the second closest friend, and so on.

4.1.2. Expectations

In addition to the primary outcomes listed in Section 5 of the primary PAP, we include a number of questions in this survey to measure potential moderators of community spillovers, including the nature of the relationship between the respondent and their friend (extended family, neighbor, co-worker, close friend), the frequency of conversation between the respondent and their friend, whether the friend was aware of the audio screening, and whether the friend was aware of the audio screening's message. We expect larger spillover effects among those friends with whom respondents engage more frequently, and among friends of those respondent's who attended the audio screening.

4.2. Children Spillover Survey

In January-February 2021, we will field a separate survey to measure spillover effects to children in the respondent's household.

4.2.1. Sampling

During the primary endline survey in December 2020, enumerators asked respondents to give the names of their children between the ages of 13 and 18. We excluded children who:

1. Were less than 13 or more than 18 years of age
2. Did not live inside the respondent's household
3. Were not biological children of the respondent

We plan to interview all children in the household who are not excluded by the criteria listed above.

During the 2021 survey, the mobilization team approached each respondent to obtain parental consent and arrange a time for the interviews. We will interview children in the household in a random order.

4.2.2. Expectations

In addition to the primary outcomes listed in Section 5, we include a number of relevant questions in this survey to facilitate the analysis of spillover effects. First, we ask children about their relationship with their parents, division of responsibilities in the household, the frequency of discussion with parents about community issues, whether they discussed the audio screening with their parents, and whether they attended the screening themselves. We expect relationship style and the frequency and depth of interactions to moderate spillovers to children; large effects are expected among those with whom the respondent engages in discussion about relevant topics. We expect spillovers to be primarily concentrated among children of individuals who attended the audio screening.

5. Primary Hypotheses

Our outcomes for the long-term radio screening experiment are drawn from the endline survey conducted one year and three months after the village screening.

The outcomes of interest are listed here for the relevant samples and with reference to the two shows, Tamapendo (T) and Wahapahapa (W).

	Endline	**Children Spillover**	**Community Spillover**
Compliance	**yes**	**yes**	**yes**
Knowledge: ● W: Increased knowledge of HIV treatment and transmission	**yes**	**yes**	**yes**
Attitudes ● T: Reduced support for early and forced marriage ● W: Increased support for HIV disclosure; reduced stigma against HIV positive individuals	**yes**	**yes**	**yes**
Perceived norms ● T: Increased perception of community/parent rejection of early marriage; increased perception of community's willingness to report EM; increased perception of community's/parent's rejection of forced marriage and prioritization of ending EM; increased perception of community's/parent's support for gender equality ● W: Decreased perception of HIV stigma in community/among parents	**yes (only community norms**	**yes (only parental norms)**	**yes (only community norms)**
Behavioral Intentions ● T: Increased willingness to report instances of EM to authorities ● W: Increased willingness to reveal HIV/AIDS status to friends/family	**yes**	**no (EM); yes (HIV)**	**yes**
Behavioral outcomes ● T: Increased willingness to speak out (and record a message) about early marriage; increased willingness to share this message with local leaders and media	**yes**	**no**	**yes**
Political priorities T*:* Votes for hypothetical candidate committed to reducing early forced marriage; increased support for women's political participation, measured in terms of both attitudes and behavioral intentions; increased priority ranking of "early forced marriage" among nine policy objectives for the village ● W*:* Votes for hypothetical candidate committed to improving access to HIV/AIDS treatment; increased priority ranking of "increasing availability of HIV/AIDS medication" among nine policy objectives for the village	**yes**	**yes (election); no (priority ranking)**	**yes**
Gender Equality ● T: Increased support for gender equality ● T: Increased rejection of intimate partner violence (IPV) ▪ T: Increased perception of community rejection of intimate partner violence	**yes**	**yes (gender equality)** **no (IPV)**	**yes**

▪ T: Increased willingness to report intimate partner violence			

Note that for some of the outcomes (rejection of EM, reporting EM, discrimination based on HIV) we also collect each partner's perception of the other partner's stance on the matter.

5.1. Survey Experiments

Within this data collection wave, we embedded a number of survey experiments.

5.1.1. Pluralistic Ignorance (only in community spillover survey)

The pluralistic ignorance survey experiment seeks to understand the role of the mismatch between first- and second- order beliefs of the respondents around Early Marriage. The pre-analysis plan for this survey experiment, including hypotheses, is contained in the Pangani FM Natural Experiment pre-analysis plan. We will implement the pluralistic ignorance survey experiment in the friends survey.

5.1.2. Laws and Persuasion (only in the community spillover survey)

This survey experiment aims to explore the effects of knowledge about laws regarding early marriage (EM) on respondents' attitudes and behaviors. It is unclear whether revisions to EM laws cause citizens to change their attitudes and behaviors, or whether citizens remain anchored to their prior beliefs, community traditions, and elite opinions even after learning about EM laws. To determine whether citizens respond to information about formal laws promulgated by Tanzania's legal institutions, we randomly assign respondents to one of three treatments:

1. Treatment 1: One-sided information about the Tanzanian Court of Appeals' recent anti-EM ruling. Respondents are informed that the Court's ruled in 2019 that it was a legal offense for a girl to marry under the age of 18, even if the girl's religion or culture allow it.
2. Treatment 2: Two-sided information about the Tanzanian Court of Appeals' recent anti-EM ruling. Respondents are informed that (1) the Tanzanian High Court prohibited marriage under the age of 18 in 2016, (2) the Attorney General challenged the ruling on the grounds that there should be a religious exemption, and (3) the Court of Appeals upheld the ruling that marriage under the age of 18 was illegal, even if the girl's religion or culture allow it.
3. Control: No information is shared.[2]

The outcomes of interest focus on the acceptability of early marriage. We are also interested in the effect of the treatment on perceived acceptability of early marriage in the community. Also varied was the cost to the respondent (the level of observability) of taking action to signal their rejection of early marriage. This behavioral measure consists of asking respondents whether they are willing to record an audio message that expresses their position on Early Marriage, either for or against, and then whether they would like to add their name to it, and whether they would like to share it with the District Council of the District and/or Pangani FM, a local radio station.

We will examine the treatment effect of each of the two treatments independently through a simple differences-in-means (OLS) as well as a covariate-adjusted regression, using lasso regression to select the minimal number of covariates that best predict each outcome. We predict positive effects on anti-EM attitudes and behaviors for the one-sided message treatment that are attenuated by the two-message treatment; however, it is unclear whether the two-message treatment leads to greater or lesser support compared to the control condition. Moreover, we are agnostic about whether the effect will be the same across attitudinal and behavioral outcomes.

5.1.3. Women's Economic Participation (only in children spillover survey)

This survey experiment aims to explore the effects of role models in the media on youth attitudes and norms around masculinity and constraints on women's economic participation. Opposition to women earning money (and especially more money than their husbands) is widespread in coastal Tanzania. We are interested in the degree to which these attitudes and norms are shaped by representations of successful marriages in mass media.

During a mid-survey refreshment break, enumerators will play one of three radio clips, each about 3 minutes long and based on real-life talk radio programs hosted on Pangani FM, a popular local radio station in Tanga Region.

1. Treatment 1: A talk radio program in which men share their experiences as husbands. In one clip, a man discusses how his relationship with his wife has improved after she started her own business. In another clip, a man discusses how his relationship was strengthened after he began sharing responsibility for household chores.

[2] Respondents in the control group are informed of the decision later in the survey

2. Treatment 2: A talk radio program in which men share their experiences as husbands. In one clip, a man discusses how his relationship with his wife deteriorated after she started her own business: she became disrespectful, and he lost his position as head of household. In another clip, a man discusses how doing household work reduces his status as a man.
3. Control condition: A music program.

The outcomes of interest are attitudes and norms about appropriate gender roles and women's economic participation. Respondent's are asked about whether it is acceptable for a woman to earn as much money as her husband and who should be responsible for various household chores. Respondent's are also asked about future behavioral intentions: whether they will pursue a career outside of the household (women) or support their wife in doing so (men).

We will examine the treatment effect of each of the two treatments independently through a simple differences-in-means (OLS) as well as covariate-adjusted regression, using LASSO to select the minimal number of covariates that best predict each outcome. We predict that we will see positive effects on pro-female economic empowerment attitudes and norms. We will examine heterogeneous treatment effects by gender, religion, and the gendered division of labor in the child's household (as reported by his or her parents in the 2020 survey). We will also conduct an automated search for heterogeneous effects.

5.1.4. Development News (only in community spillover survey)

This survey experiment explores the first-order effects of development-related news reporting on attitudes about local government effectiveness and trustworthiness of campaign promises, as well as the second-order effects on voting behavior and responsibility attribution.

Respondents are randomly assigned to one of two treatments or a control condition. The treatments are novel radio segments in which the radio host of Pangani FM introduces a development-related hard news story and plays real-life clips of villagers speaking about development issues.

- Treatment 1: "Positive" news story. The news story highlights instances in which candidates for local office made campaign promises about development projects and successfully delivered them after the election. It features one villager speaking about a campaign promise to improve water access and another villager speaking about a promise to improve roads, both of which were successfully implemented.

- Treatment 2: "Negative" news story. The news story highlights instances in which candidates for local office made campaign promises about development projects but failed to deliver them after the election. It features one villager speaking about a campaign promise to improve roads and another villager speaking about a promise to improve schools, both of which were not subsequently implemented.

- Pure control.

The outcomes of interest are attitudes about the effectiveness of local government in administering development, attitudes about the trustworthiness of development-related campaign promises, attitudes about which actors are most responsible for development, and candidate choice in a hypothetical election. For the latter outcome, respondents are given a choice between a candidate with "bad" personal attributes but a highly relevant campaign promise (Candidate 1) or a candidate with "good" personal attributes but a less relevant campaign promise (Candidate 2). Campaign promises depend on each respondent's unique ranking of none political priorities, as measured earlier in the political priorities section of the survey. For each respondent, the distance between the relevance of the election promises is randomly varied (i.e. first priority vs. ninth priority; second priority vs. eighth priority; third priority vs. seventh priority; and fourth priority vs. sixth priority). Random variation in the distance between election promises allows for the generation of comparative statics for the effect of candidate policy distance on issue voting (as opposed to personal or non-issue voting). We hypothesize that the greater the distance, the more likely respondents will be to issue vote.

We hypothesize that the negative (positive) treatment will decrease (increase) perceptions of local government effectiveness, decrease (increase) perceptions of the trustworthiness of development-related campaign promises, and decrease (increase) support for Candidate 1 (increase support for Candidate 2). We are agnostic about the effect of the treatments on perceptions of who is responsible for development (local government, national government, or villagers themselves) and conduct a two-sided test for this outcome.

We will examine the effect of each treatment with respect to the control group as well as with respect to the other treatment using a simple difference-in-means estimator (OLS) as well as covariate-adjusted regression, using LASSO to select the minimal number of covariates that best predict each outcome. We will test for an interaction between the treatment and the issue distance of the candidates. We will also conduct an automated search for heterogeneous effects.

6. Estimation

6.1. General Coding Rules

- Unless otherwise indicated, outcome measures will be constructed from single items
- We will construct variables so that higher values represent more agreement with the norms portrayed in the radio screenings.
- Where more than one survey response relates to a common outcome construct, we will generate an index in the following manner: all variables will be standardized to range from 0 to 1 by dividing by the maximum possible value. All variables will then be summed and divided by the number of variables in the index.

6.2. Compliance

We expect that the effects of screening will be larger for baseline respondents who attended the screening, as compared to noncompliers who did not attend. The placebo design allows us to compare just compliers in both experimental groups. The midline results confirmed that compliance is not significantly related to assigned treatment; we will check to make sure that this finding persists in the endline given that some attrition will occur in our survey panel. We will also check balance to verify that compliers' partners have similar profiles in both experimental groups.

6.3. Primary Specification

For all analyses, we will estimate treatment effects using least squares regression and clustered standard errors, with fixed effects for block (ward). Additional covariates will be selected using the process outlined in Section 5.6.

For example, restricting the sample to compliers' partners, we would assess the (spillover) effects of the treatment on partners' attitudes about forced marriage by regressing this outcome on a treatment indicator and block indicators.

6.4. Basic Specification

For all analyses, we will also estimate treatment effects using only fixed effects for block (ward) and no other covariates.

6.5. Randomization Inference

For all specifications, we will calculate p-values with randomization inference using the randomization code provided in Appendix 2. We will calculate cluster-robust standard errors using the sandwich package for R but will not rely on these standard errors for inference.

6.6. Covariates

As noted above, we will include an indicator for block (ward) in all specifications.

We will use lasso to select the minimal number of covariates that best predict each outcome and include only these covariates in our estimation. In our "agnostic" specification, we will draw potential covariates exclusively from the baseline survey. In our "extensive" specification, we include covariates measured in the endline that, although posttreatment, are not plausibly affected by our interventions. These include measures of political interest, political participation, and psychological traits identified in the literature as promoting narrative persuasion (openness, need for cognition, and absorption).

We will perform lasso on covariate data that has been pre-processed to remove item-level missingness through multiple chain equations, as described below. The lasso procedure that we plan to use features a generalized linear model with lasso penalization and is implemented in the glmnet package for R. The loss function requires selecting a regularization parameter that determines the severity of the penalty for including extra covariates. Since this regularization parameter cannot be optimally chosen in advance, we will select it using 10-fold cross-validation. Specifically, for each outcome, we will choose the regularization parameter that minimizes the 10-fold cross-validation error averaged over 10 runs (since the folds are chosen at random). Only the covariates retained by the lasso will be included in the specification.

We will perform this lasso variable selection method using the entire list of baseline covariates, including block and resample indicators, among the entire sample. For extensive specification, the process will be repeated for the expanded set of covariates. Note that this procedure implies that for each outcome, a different number of covariates may be included on the right-hand side.

6.7. Heterogeneous Treatment Effects

We will explore individual-level treatment heterogeneity using the generalized causal forest algorithm in R, as we have done for the midline results. When the sources of treatment effect heterogeneity are not known *ex ante*, restricting analyses to pre-registered hypotheses may hinder the discovery of unanticipated results; however, ad hoc exploration of treatment effect heterogeneity runs the risk of false discovery. We therefore use machine learning tools originally developed to predict individualized responses to medical interventions in order to automate and synthesize the search for treatment effect heterogeneity across many covariates.

For the purpose of this analysis we will use the large set of covariates collected at baseline, as well as other post-treatment covariates that are plausibly unaffected by treatment status. These plausibly exogenous attributes include parenting style, partner relationship measures (especially discussion), levels of political interest and participation, and psychological traits.

6.8. Multiple Comparisons

We address the problem of multiple comparisons in the following way. Estimated t-ratios for the analyses will be sorted in descending order for all one-tailed hypothesis. A null distribution of t-ratios will be simulated by repeating the random assignment 10,000 times and for each simulation generating a vector of estimated t-ratios. To obtain the p-value for the largest estimated t-ratio, we will ask: what proportion of the simulations generated a largest p-value at least as large as what we obtained empirically? The same procedure will be applied to the second largest p-value, and so forth (Anderson 2008).

6.9. Item Level Missingness

Non-response to outcome questions will be dealt with through multivariate imputation via chained equations (MICE) using the mice package for R. Imputations for outcome measures will be performed within issue areas (IPV, HIV, EM).

F Measures

This section documents the study's outcome measures. Table A20 reports each unique survey item once: its full text, response scale and coding, the respondent groups to whom it was administered, and its source, distinguishing items adapted from the Demographic and Health Surveys AIDS module, the UNAIDS discriminatory-attitudes indicators, and published stigma scales from items developed for this study. Table A21 reports index composition, measurement timing, and internal consistency for every index by sample. Because the indices are formative summary measures rather than scales for a single latent construct, internal consistency varies by design: it is high where items tap a common attitude (disclosure: Cronbach's α = 0.77–0.86), moderate for expressed stigma (α = 0.42–0.46), and low for the knowledge and prioritization indices (α = 0.07–0.27), whose components are distinct facts and behaviors that the intervention could move independently.

Table A20: Outcome Measures: Item Text, Scales, Coding, Administration, and Sources

Item	Item Text	Response Scale and Coding	Asked of	Source
HIV Knowledge				
HIV Transmit	To the best of your understanding, can a person with HIV/AIDS transmit the disease to the people that he or she works with?	Yes/No; 1 = HIV cannot be transmitted to coworkers, 0 = otherwise (reverse-coded)	R, P	Adapted from the casual-contact misconception items of the DHS AIDS module (Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC) [Tanzania Mainland], Ministry of Health (MoH) [Zanzibar], National Bureau of Statistics (NBS), Office of the Chief Government Statistician (OCGS), and ICF 2016)
ARV Survive	In your understanding, what are the most important things a person with HIV/AIDS can do to help live longer?	Open-ended, enumerator-coded; 1 = mentioned ARVs, 0 = otherwise	R, P, F, C	Adapted from ART knowledge items in Tanzanian national HIV surveys (National Ministry of Health, Community Development, Gender, Elderly, and Children 2017); open-ended format adapted from the research team's prior survey instruments [REDACTED for review]

continued on next page

(continued from previous page)

Item	Item Text	Response Scale and Coding	Asked of	Source
ARV No Spread	In your understanding, what are the most important things a person with HIV/AIDS can do to make sure that they do not infect their partner with HIV?	Open-ended, enumerator-coded; 1 = mentioned ARVs, 0 = otherwise	R, P	Adapted from ART knowledge items in Tanzanian national HIV surveys (National Ministry of Health, Community Development, Gender, Elderly, and Children 2017); open-ended format adapted from the research team's prior survey instruments [REDACTED for review]
ARV Prompted	Have you heard about antiretroviral drugs or ARVs that people infected with HIV can get from a doctor or a nurse to help them live longer?	Yes/No; 1 = heard of ARVs, 0 = otherwise	R, P, F	Adapted from ART awareness items in Tanzanian national HIV surveys (National Ministry of Health, Community Development, Gender, Elderly, and Children 2017)
Willingness to Disclose				
No Secret	If a member of your family got HIV, would you want it to remain a secret?	Yes/No; 1 = would not want it to remain a secret, 0 = otherwise (reverse-coded)	R, P, C	Adapted from the DHS AIDS module; a component of the UNAIDS shame-and-stigma indicator set (Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC) [Tanzania Mainland], Ministry of Health (MoH) [Zanzibar], National Bureau of Statistics (NBS), Office of the Chief Government Statistician (OCGS), and ICF 2016; UNAIDS 2022)

continued on next page

(continued from previous page)

Item	Item Text	Response Scale and Coding	Asked of	Source
Disclose to Coworker	If you had HIV, would you tell or not tell your coworkers?	Tell/Not tell; 1 = would tell coworkers, 0 = otherwise	R, P, F	Adapted from the research team's prior survey instruments [REDACTED for review]
Disclose to Friend	If you had HIV, would you tell or not tell your friends?	Tell/Not tell; 1 = would tell friends, 0 = otherwise	R, P, F	Adapted from the research team's prior survey instruments [REDACTED for review]
Expressed Stigma				
Sit on Bus	Would you be willing to sit next to someone with HIV/AIDS on the bus?	Yes/No; 1 = willing to sit next to, 0 = otherwise	R, P, F, C	Adapted from social-distance stigma items validated in southern and eastern Africa (Kalichman et al. 2005)
Not Fired	In your opinion, if a field worker on a large farm becomes infected with HIV, should they be fired or should they be allowed to keep their job?	Fired/Keep job; 1 = allowed to keep job, 0 = fired (reverse-coded)	R, P	Adapted from the UNAIDS/DHS discriminatory-attitudes item on whether a person with HIV should be allowed to continue working (UNAIDS 2022; Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC) [Tanzania Mainland], Ministry of Health (MoH) [Zanzibar], National Bureau of Statistics (NBS), Office of the Chief Government Statistician (OCGS), and ICF 2016)
Not Fired Norm	What do you think others in your community think about this issue? [asked following the preceding item]	1 = believes community thinks the worker should keep their job, 0 = otherwise	R, P	Perceived-community-norm variant of the preceding item, adapted from the research team's prior survey instruments [REDACTED for review]

continued on next page

(continued from previous page)

Item	Item Text	Response Scale and Coding	Asked of	Source
Parents' Advice for Bus	Would your parents advise you not to sit next to someone who has HIV on the bus? Or would they think it is ok?	Yes/No; 1 = parents would not advise against sitting, 0 = otherwise (reverse-coded)	C	Child-specific variant of the social-distance item, adapted from the research team's prior survey instruments [REDACTED for review]
HIV Policy Prioritization				
Elect	The first candidate is named (name), is a (Muslim or Christian), and promises villagers to make HIV/AIDS treatment more available in the village. The second candidate is named (name), is a (Muslim or Christian), and promises villagers to (improve roads / reduce crime). Which of these candidates do you think should be selected?	Binary candidate choice, randomized platform and attribute order; 1 = chose the HIV-platform candidate, 0 = otherwise	R, P, F, C	Candidate-choice vignette adapted from the research team's prior survey instruments [REDACTED for review]
HIV First	Here is another set of cards, which show different goals for your village. Please put them in order, from most important to you to least important to you. [nine goals, including increasing access to health services such as HIV/AIDS medication]	Card ranking of nine village goals; 1 = ranked HIV/AIDS medication access first, 0 = otherwise	R, P, F	Card-ranking priority measure adapted from the research team's prior survey instruments [REDACTED for review]

Note: One row per unique survey item, grouped by outcome family. *Asked of*: R = primary respondents, P = partners, F = friends, C = teenage children. Respondents and partners were surveyed at the 18-month endline; friends and children were surveyed two to three months later (20–21 months post-treatment). All items were administered in Swahili. Items marked "reverse-coded" are recoded so that higher values indicate the outcome named in the item label. Sources marked [REDACTED for review] are the survey instruments of the research team's prior evaluations (the two-week midline evaluation of this intervention and a prior radio-drama field experiment), withheld for anonymous review. Index composition and internal consistency by sample appear in Table A21.

Table A21: **Index Composition, Measurement Timing, and Internal Consistency**

Index	Sample	Measurement point	Items	Cronbach's α	N
Knowledge	Respondents	18-month endline	4	0.13	1004
Stigma	Respondents	18-month endline	3	0.42	1012
Disclosure	Respondents	18-month endline	3	0.77	1013
Priority	Respondents	18-month endline	2	0.13	987
Knowledge	Partners	Endline wave	4	0.27	703
Stigma	Partners	Endline wave	3	0.46	698
Disclosure	Partners	Endline wave	3	0.77	707
Priority	Partners	Endline wave	2	0.07	709
Knowledge	Friends	Endline +2–3 months	2	-0.04	957
Disclosure	Friends	Endline +2–3 months	2	0.86	970
Priority	Friends	Endline +2–3 months	2	0.19	971

Single-item outcomes (friend stigma; all teenage-child outcomes, measured at endline +2–3 months) have no internal-consistency statistic.

Note: Cronbach's α computed on complete cases for each index and sample. The endline wave ran 18 months post-treatment; friend and child surveys followed two to three months later. Indices are equally weighted means of available items, constructed as pre-specified summary indices (Kling et al. 2007) rather than reflective scales of a latent construct; low internal consistency for the knowledge and prioritization indices reflects that their components are distinct facts and behaviors. Item-level treatment effects appear alongside every index in the results tables.

G Attrition and Compliance by Treatment Arm

Table A22: **Attrition and Compliance**

	Compliance	**Attrition**		**Spillover Sampling**		
	Compliance	Midline	Endline	Partner	Friend	Children
Treatment Effect	0.006	0.003	0.003	0.043	0.032	0.012
Standard Error	(0.020)	(0.005)	(0.010)	(0.025)	(0.020)	(0.042)
RI p-value	0.857	0.812	0.896	0.268	0.353	0.814
Test Direction	Two	Two	Two	Two	Two	Two
Placebo Mean	0.87	0.96	0.96	0.66	0.9	0.69
Placebo Village SD	0.12	0.02	0.04	0.17	0.11	0.25
DV Range	[1-0]	[1-0]	[1-0]	[1-0]	[1-0]	[6-0]
Blocked FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00	0.00	0.00
Observations	1,205	1,205	1,205	1,205	1,205	1,205

Note: * p < 0.05. Randomization *p*-values are calculated with 2,000 randomizations testing the null hypothesis of no effect for any observation. Column 1 was coded as a 1 if respondents attended the screening and 0 if otherwise. Column 2 was coded as a 1 if respondents were interviewed at midline and 0 if otherwise. Column 3 was coded as a 1 if respondents were interviewed at endline and 0 if otherwise. Column 4 was coded as a 1 if respondents had a partner who was interviewed and 0 if otherwise. Column 5 was coded as a 1 if respondents had a friend who was interviewed and 0 if otherwise. Column 6 counts the number of the respondent's teenage children who were interviewed (0 to 6).

H Effect Heterogeneity

We report exploratory (not pre-registered) analyses of treatment-effect heterogeneity among primary respondents across four standard demographic dimensions: gender, religion, education, and wealth (proxied by housing quality). For each moderator, we report the treatment effect within each subgroup (ordinary least squares with randomization-block fixed effects and standard errors clustered by village), the treatment-by-moderator interaction from the pooled model, and the two-sided p-value on the interaction. Across the sixteen interaction tests, none reaches conventional significance; there is very little evidence of heterogeneity on standard demographic variables.

H.1 Gender

Table A23: **Treatment Effect Heterogeneity by Respondent Gender**

	Women		Men				
	Effect	SE	Effect	SE	Interaction	p (interaction)	N
Knowledge Index	0.014	(0.010)	0.030	(0.009)	-0.014	0.446	1018
Stigma Index	0.008	(0.010)	0.014	(0.014)	-0.007	0.773	1018
Disclosure Index	0.067	(0.033)	0.088	(0.032)	-0.019	0.758	1018
Priority Index	0.046	(0.019)	0.070	(0.021)	-0.019	0.657	1019

Note: Subgroup columns report the treatment effect estimated within each subgroup; the interaction column reports the treatment × moderator coefficient from the pooled model with randomization-block fixed effects and standard errors clustered by village. p-values are two-sided. These analyses are exploratory and were not pre-registered.

H.2 Religion

Table A24: **Treatment Effect Heterogeneity by Religion**

	Muslim		Non-Muslim				
	Effect	SE	Effect	SE	Interaction	p (interaction)	N
Knowledge Index	0.028	(0.009)	0.013	(0.009)	0.017	0.372	1018
Stigma Index	0.007	(0.012)	0.027	(0.012)	-0.014	0.445	1018
Disclosure Index	0.078	(0.030)	0.094	(0.036)	-0.003	0.956	1018
Priority Index	0.067	(0.019)	0.055	(0.029)	0.022	0.622	1019

Note: Subgroup columns report the treatment effect estimated within each subgroup; the interaction column reports the treatment × moderator coefficient from the pooled model with randomization-block fixed effects and standard errors clustered by village. p-values are two-sided. These analyses are exploratory and were not pre-registered.

H.3 Education

Table A25: **Treatment Effect Heterogeneity by Education (Reached Standard 7)**

	Standard 7+		Below Standard 7				
	Effect	SE	Effect	SE	Interaction	p (interaction)	N
Knowledge Index	0.027	(0.008)	0.011	(0.020)	0.020	0.420	998
Stigma Index	0.017	(0.008)	-0.006	(0.021)	0.026	0.350	998
Disclosure Index	0.064	(0.025)	0.138	(0.034)	-0.054	0.298	998
Priority Index	0.064	(0.014)	0.037	(0.045)	0.020	0.738	999

Note: Subgroup columns report the treatment effect estimated within each subgroup; the interaction column reports the treatment × moderator coefficient from the pooled model with randomization-block fixed effects and standard errors clustered by village. p-values are two-sided. These analyses are exploratory and were not pre-registered.

H.4 Wealth

Table A26: **Treatment Effect Heterogeneity by Housing Quality (Mud Walls)**

	Mud Walls		Non-Mud Walls				
	Effect	SE	Effect	SE	Interaction	p (interaction)	N
Knowledge Index	0.007	(0.010)	0.056	(0.017)	-0.055	0.054	1018
Stigma Index	0.008	(0.010)	0.028	(0.014)	-0.016	0.435	1018
Disclosure Index	0.070	(0.028)	0.034	(0.050)	-0.004	0.954	1018
Priority Index	0.064	(0.016)	0.013	(0.035)	0.036	0.446	1019

Note: Housing quality (whether the dwelling has mud walls) proxies household wealth. Subgroup columns report the treatment effect estimated within each subgroup; the interaction column reports the treatment × moderator coefficient from the pooled model with randomization-block fixed effects and standard errors clustered by village. p-values are two-sided. These analyses are exploratory and were not pre-registered.

I Statistical Power

As noted in the main text, the sample size was chosen to balance budget constraints and a minimum detectable effect target of roughly 0.2 standard deviations on primary outcomes, although the unknown intracluster correlation precluded a definitive ex ante power calculation. Table A27 characterizes the statistical power of the realized design: for each index outcome, it reports the minimum detectable effect (MDE) at 80% power and $\alpha = .05$ under the pre-registered one-sided test, computed as 2.49 times the cluster-robust standard error, in raw units and in placebo-group standard deviations. Because these standard errors are clustered by assignment village, within-village dependence is reflected in the MDE empirically rather than through an assumed design effect. The table also reports, for each outcome, the unconditional village-level intraclass correlation (ICC), the clustering statistic that governs power in a cluster-randomized design (through the design effect $1 + (\bar{m} - 1)\,\text{ICC}$, where $\bar{m}$ is the average cluster size): raw village ICCs range from near zero to 0.08, so with roughly 34 respondents per village the clustering materially reduces effective sample size, which the matched-pair blocking then largely absorbs (the ICC conditional on blocking is near zero; see Appendix M). The design affords 80% power to detect one-sided effects of roughly 0.10–0.16 placebo standard deviations for primary respondents and 0.12–0.22 for the spillover samples. Observed treatment effects on the primary indices lie near this one-sided detection boundary the study was designed around.

Table A27: **Minimum Detectable Effects, Index Outcomes**

			MDE (one-sided)		
	SE	ICC	Raw	Placebo SD	N
Panel A: Endline Respondents					
Knowledge Index	0.007	0.005	0.018	0.11	1018
Stigma Index	0.009	0.030	0.023	0.15	1018
Disclosure Index	0.024	0.040	0.060	0.15	1018
Priority Index	0.013	0.025	0.033	0.10	1019
Panel B: Partners					
Knowledge Index	0.010	0.015	0.025	0.13	709
Stigma Index	0.015	0.037	0.036	0.16	709
Disclosure Index	0.025	0.048	0.063	0.16	709
Priority Index	0.021	0.052	0.053	0.18	709
Panel C: Friends					
Knowledge Index	0.028	0.077	0.070	0.22	972
Stigma (Sit on Bus)	0.018	0.051	0.046	0.14	972
Disclosure Index	0.022	0.015	0.055	0.12	972
Priority Index	0.017	0.028	0.042	0.13	972
Panel D: Teenage Children					
Knowledge (ARV Open-Ended)	0.028	0.027	0.071	0.20	496
Disclosure (Family Keep Secret)	0.035	0.049	0.088	0.18	495
Stigma (Sit on Bus)	0.041	0.068	0.103	0.22	496
Priority (Elect)	0.024	0.000	0.061	0.13	496

Note: MDEs at 80% power and $\alpha = .05$ under the pre-registered one-sided test, equal to $(z_{1-\alpha} + z_{0.80}) \times SE$ = 2.49 × SE, where SE is the cluster-robust standard error (randomization-block fixed effects, clustering by assignment village). The Placebo SD column expresses the MDE in placebo-group individual-level standard deviations. ICC is the unconditional village-level intraclass correlation for each outcome, estimated from a null variance-components model; it is the clustering statistic relevant to power via the design effect $1 + (\bar{m} - 1)\,\text{ICC}$. The ICC conditional on the matched-pair blocking is near zero (Appendix M).

J Robustness

The appendix balance table (Table A4) shows a nominal treatment–placebo difference in household television ownership; of 50 baseline covariates, one (2.0%) shows a difference exceeding the 5% randomization threshold and two (4.0%) the 10% threshold, fewer than chance would predict. As a direct check that this imbalance does not drive our results, Table A28 re-estimates treatment effects on the four primary indices controlling for television ownership. Estimates are essentially unchanged.

Table A28: **Primary Indices, Adjusting for Television Ownership**

	Unadjusted		Adjusted			
	Est.	SE	Est.	SE	p (two-sided)	N
Knowledge Index	0.021	(0.007)	0.022	(0.007)	0.007	1018
Stigma Index	0.011	(0.009)	0.011	(0.009)	0.248	1018
Disclosure Index	0.073	(0.024)	0.072	(0.024)	0.006	1018
Priority Index	0.055	(0.013)	0.056	(0.013)	0.000	1019

Note: Both specifications are ordinary least squares with randomization-block fixed effects and standard errors clustered by assignment village; the adjusted specification adds an indicator for baseline household television ownership. p-values are two-sided (analytic).

K Standardized Effect Sizes

Table A29: **Standardized Effect Sizes, All Outcomes**

	d	95% CI	*N*
Panel A: Endline Respondents			
Knowledge Index	0.14	[0.05, 0.23]	1018
Knowledge: Transmit	0.01	[-0.07, 0.08]	1010
Knowledge: ARV Survive	0.08	[-0.00, 0.16]	1018
Knowledge: ARV No Spread	0.14	[0.05, 0.23]	1018
Knowledge: ARV Prompted	0.07	[0.00, 0.13]	1012
Stigma Index	0.07	[-0.05, 0.19]	1018
Stigma: Sit on Bus	0.02	[-0.05, 0.09]	1016
Stigma: Not Fired	0.02	[-0.10, 0.14]	1016
Stigma: Not Fired Norm	0.08	[-0.02, 0.19]	1014
Disclosure Index	0.18	[0.06, 0.29]	1018
Disclosure: No Secret	0.14	[0.04, 0.25]	1018
Disclosure: Coworker	0.15	[0.04, 0.26]	1015
Disclosure: Friend	0.16	[0.05, 0.27]	1015
Priority Index	0.17	[0.09, 0.25]	1019
Priority: Elect	0.17	[0.08, 0.25]	1019
Priority: HIV First	0.09	[-0.02, 0.20]	987
Panel B: Partners			
Knowledge Index	0.05	[-0.05, 0.15]	709
Stigma Index	-0.02	[-0.15, 0.11]	709
Disclosure Index	0.16	[0.04, 0.29]	709
Priority Index	0.17	[0.03, 0.31]	709
Panel C: Friends			
Knowledge Index	0.01	[-0.16, 0.19]	972
Stigma: Sit on Bus	-0.19	[-0.30, -0.08]	972
Disclosure Index	0.02	[-0.07, 0.12]	972
Priority Index	0.05	[-0.06, 0.15]	972
Panel D: Teenage Children			
Knowledge: ARV Open-Ended	-0.04	[-0.19, 0.12]	496
Disclosure: Family Keep Secret	0.23	[0.09, 0.37]	495
Stigma: Sit on Bus	-0.01	[-0.18, 0.16]	496
Stigma: Parents Advice	-0.03	[-0.18, 0.12]	495
Priority: Elect	-0.04	[-0.14, 0.06]	496

Note: Standardized effect sizes following the logic of Morris (2008): the covariate-adjusted treatment effect (ordinary least squares with randomization-block fixed effects) divided by the placebo-group individual-level standard deviation, with Hedges' small-sample correction. Because most outcomes were not measured at baseline, the posttest-only variant is reported. 95% confidence intervals are based on standard errors clustered by assignment village.

L Design-Based Estimates

The tables in this section report the design-based counterpart to the multilevel mixed models presented in the main text: ordinary least squares with randomization-block fixed effects and standard errors clustered by assignment village, with the same pre-registered one-tailed randomization-inference p-values (2,000 permutations of village-level treatment assignment within matched ward pairs). Beneath each pre-registered directional test, the tables also report the exact two-sided randomization-inference p-value (the share of permuted estimates at least as large in absolute value as the observed estimate, from the same permutations), so that readers may apply either standard. Because the randomization-block fixed effects absorb the two village means within each matched pair, point estimates coincide with the mixed-model estimates up to rounding, and the randomization-inference p-values agree except where an estimate rounds differently. Variable coding for each outcome is described in the notes to the corresponding main-text tables.

Table A30: **Effect of HIV Drama on HIV Knowledge, Design-Based Estimates**

	Index	Transmit	ARV Survive	ARV No Spread	ARV Prompted
Treatment Effect	0.021	0.001	0.03	0.027	0.025
Standard Error	(0.007)	(0.008)	(0.016)	(0.009)	(0.012)
RI p-value	0.033	0.466	0.108	0.033	0.098
RI p-value (two-sided)	0.068	0.942	0.208	0.062	0.188
Test Direction	(+)	(+)	(+)	(+)	(+)
Placebo Mean	0.66	0.96	0.83	0.04	0.83
Placebo Village SD	0.03	0.03	0.07	0.04	0.07
DV Range	[1-0]	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00	0.00
Observations	1,018	1,010	1,018	1,018	1,012

Note: Ordinary least squares with randomization-block fixed effects; standard errors clustered by assignment village. One-tailed randomization inference p-values are calculated with 2,000 permutations of village-level treatment assignment; the two-sided row reports the exact two-sided p-value from the same permutations. Variable coding follows Table 3.

Table A31: **Effect of HIV Drama on HIV Disclosure, Design-Based Estimates**

	Index	No Secret	Coworker	Friend
Treatment Effect	0.073	0.069	0.075	0.08
Standard Error	(0.024)	(0.025)	(0.028)	(0.028)
RI p-value	0.033	0.04	0.044	0.034
RI p-value (two-sided)	0.061	0.079	0.085	0.069
Test Direction	(+)	(+)	(+)	(+)
Placebo Mean	0.57	0.64	0.49	0.58
Placebo Village SD	0.11	0.1	0.13	0.13
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00
Observations	1,018	1,018	1,015	1,015

Note: Ordinary least squares with randomization-block fixed effects; standard errors clustered by assignment village. One-tailed randomization inference p-values are calculated with 2,000 permutations of village-level treatment assignment; the two-sided row reports the exact two-sided p-value from the same permutations. Variable coding follows Table 4.

Table A32: **Effect of HIV Drama on HIV Stigma, Design-Based Estimates**

	Index	Sit on Bus	Not Fired	Not Fired Norm
Treatment Effect	0.011	0.003	0.003	0.026
Standard Error	(0.009)	(0.006)	(0.011)	(0.017)
RI p-value	0.186	0.356	0.41	0.133
RI p-value (two-sided)	0.41	0.751	0.853	0.288
Test Direction	(+)	(+)	(+)	(+)
Placebo Mean	0.94	0.97	0.97	0.89
Placebo Village SD	0.03	0.03	0.04	0.06
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00
Observations	1,018	1,016	1,016	1,014

Note: Ordinary least squares with randomization-block fixed effects; standard errors clustered by assignment village. One-tailed randomization inference p-values are calculated with 2,000 permutations of village-level treatment assignment; the two-sided row reports the exact two-sided p-value from the same permutations. Variable coding follows Table 5.

Table A33: **Effect of HIV Drama on HIV Priorities, Design-Based Estimates**

	Index	Elect	HIV First
Treatment Effect	0.055	0.084	0.031
Standard Error	(0.013)	(0.022)	(0.019)
RI p-value	0.011	0.011	0.164
RI p-value (two-sided)	0.021	0.02	0.319
Test Direction	(+)	(+)	(+)
Placebo Mean	0.33	0.51	0.13
Placebo Village SD	0.07	0.12	0.05
DV Range	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes
Covariates	No	No	No
Adjusted R2	0.00	0.00	0.00
Observations	1,019	1,019	987

Note: Ordinary least squares with randomization-block fixed effects; standard errors clustered by assignment village. One-tailed randomization inference p-values are calculated with 2,000 permutations of village-level treatment assignment; the two-sided row reports the exact two-sided p-value from the same permutations. Variable coding follows Table 6.

Table A34: **Spillover Effects on Partners, Design-Based Estimates**

	Knowledge	**Disclosure**	**Stigma**	**Priority**
	Index	Index	Index	Index
Treatment Effect	0.009	0.067	-0.005	0.05
Standard Error	(0.01)	(0.025)	(0.015)	(0.021)
RI p-value	0.278	0.04	0.577	0.068
RI p-value (two-sided)	0.54	0.086	0.806	0.134
Test Direction	(+)	(+)	(+)	(+)
Placebo Mean	0.64	0.54	0.9	0.32
Placebo Village SD	0.05	0.13	0.06	0.07
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00
Observations	709	709	709	709

Note: Ordinary least squares with randomization-block fixed effects; standard errors clustered by assignment village. One-tailed randomization inference p-values are calculated with 2,000 permutations of village-level treatment assignment; the two-sided row reports the exact two-sided p-value from the same permutations. Variable coding follows Table 7.

Table A35: **Spillover Effects on Friends, Design-Based Estimates**

	Knowledge	**Disclosure**	**Stigma**	**Priority**
	Index	Index	Sit on Bus	Index
Treatment Effect	0.004	0.01	-0.062	0.015
Standard Error	(0.028)	(0.022)	(0.018)	(0.017)
RI p-value	0.477	0.368	0.983	0.274
RI p-value (two-sided)	0.936	0.758	0.028	0.532
Test Direction	(+)	(+)	(+)	(+)
Placebo Mean	0.72	0.57	0.88	0.36
Placebo Village SD	0.12	0.09	0.07	0.07
DV Range	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00
Observations	972	972	972	972

Note: Ordinary least squares with randomization-block fixed effects; standard errors clustered by assignment village. One-tailed randomization inference p-values are calculated with 2,000 permutations of village-level treatment assignment; the two-sided row reports the exact two-sided p-value from the same permutations. Variable coding follows Table 8.

Table A36: **Spillover Effects on Teenage Children, Design-Based Estimates**

	Knowledge	**Disclosure**	**Stigma**		**Priority**
	ARV Open-Ended	Family Keep Secret	Sit on Bus	Parents Advice for Bus	Elect
Treatment Effect	-0.014	0.114	-0.006	-0.015	-0.018
Standard Error	(0.028)	(0.035)	(0.041)	(0.036)	(0.024)
RI p-value	0.626	0.018	0.546	0.616	0.701
RI p-value (two-sided)	0.748	0.038	0.912	0.755	0.588
Test Direction	(+)	(+)	(+)	(+)	(+)
Placebo Mean	0.85	0.52	0.67	0.66	0.67
Placebo Village SD	0.11	0.12	0.18	0.18	0.13
DV Range	[1-0]	[1-0]	[1-0]	[1-0]	[1-0]
Blocked FE	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No
Adjusted R2	0.00	0.00	0.00	0.00	0.00
Observations	496	495	496	495	496

Note: Ordinary least squares with randomization-block fixed effects; standard errors clustered by assignment village. One-tailed randomization inference p-values are calculated with 2,000 permutations of village-level treatment assignment; the two-sided row reports the exact two-sided p-value from the same permutations. Variable coding follows Table 9.

M Multilevel Mixed-Model Comparison

The main-text tables report our primary specification: linear multilevel mixed models entering randomization-block (ward-pair) fixed effects alongside a village random intercept (MLM-A), with one-tailed randomization-inference p-values. This appendix compares that specification, outcome by outcome, against the design-based blocked least-squares benchmark (Appendix L) and against a fully multilevel variant entering random intercepts for both randomization blocks and villages (MLM-B). The analytic p-values shown here use Kenward-Roger small-sample degree-of-freedom corrections, which guard against anti-conservative inference with 30 clusters, and are two-sided; they are diagnostics for comparing the specifications rather than our primary design-based tests.

Two features of the results bear note. First, point estimates are nearly identical across all three specifications, and no substantive conclusion changes. Second, estimated village-level intraclass correlations are small once randomization blocks are accounted for (at most 0.03 for primary-respondent indices and below 0.10 across all index outcomes). In a matched-pair design with two villages per block, block fixed effects absorb nearly all between-village variation; village random-intercept variances are accordingly estimated at or near zero for several outcomes (a boundary estimate in the mixed-model literature), in which case the mixed model effectively reduces to the blocked ordinary least squares specification. This pattern indicates that the interdependence of respondents within villages, conditional on the design, is minimal, and that design-based inference does not understate uncertainty relative to the multilevel alternative.

Table A37: **Design-Based versus Multilevel Mixed-Model Estimates**

	Design-based		MLM-A			MLM-B		
	Est.	SE	Est.	SE	p (KR)	p (KR)	ICC	N
Panel A: Endline Respondents								
Knowledge Index	0.021	0.007	0.022	0.010	0.057	0.060	0.000	1018
Stigma Index	0.011	0.009	0.010	0.013	0.444	0.461	0.027	1018
Disclosure Index	0.073	0.024	0.072	0.034	0.054	0.062	0.026	1018
Priority Index	0.055	0.013	0.055	0.020	0.017	0.018	0.000	1019
Panel B: Partners								
Knowledge Index	0.009	0.010	0.009	0.014	0.524	0.492	0.000	709
Stigma Index	-0.005	0.015	-0.005	0.021	0.798	0.715	0.019	709
Disclosure Index	0.067	0.025	0.066	0.036	0.089	0.094	0.018	709
Priority Index	0.050	0.021	0.048	0.029	0.126	0.158	0.026	709
Panel C: Friends								
Knowledge Index	0.004	0.028	0.003	0.039	0.944	0.957	0.080	972
Stigma (Sit on Bus)	-0.062	0.018	-0.064	0.026	0.031	0.029	0.011	972
Disclosure Index	0.010	0.022	0.010	0.032	0.756	0.805	0.005	972
Priority Index	0.015	0.017	0.014	0.024	0.556	0.630	0.011	972
Panel D: Teenage Children								
Knowledge (ARV Open-Ended)	-0.014	0.028	-0.012	0.038	0.752	0.727	0.015	496
Stigma (Sit on Bus)	-0.006	0.041	-0.007	0.059	0.913	0.884	0.049	496
Disclosure (Family Keep Secret)	0.114	0.035	0.117	0.049	0.035	0.061	0.016	495
Priority (Elect)	-0.018	0.024	-0.018	0.044	0.697	0.614	0.000	496

Note: Design-based columns report ordinary least squares with randomization-block fixed effects and standard errors clustered by village (Appendix L). MLM-A reports the main-text primary specification: linear mixed models with randomization-block fixed effects and a village random intercept; MLM-B replaces block fixed effects with a block random intercept alongside the village random intercept. Estimates and standard errors shown are from MLM-A; MLM-B estimates are nearly identical. p-values are two-sided with Kenward-Roger small-sample degree-of-freedom corrections. ICC reports the village-level intraclass correlation from MLM-B. Village random-intercept variances estimated at zero (boundary fits) are expected under a matched-pair design, in which block effects absorb nearly all between-village variation; in such cases the mixed model reduces to the blocked least-squares specification.