JAMES V. MARRONE, TODD C. HELMUS, ELIZABETH BODINE-BARON, CHRISTOPHER SANTUCCI

Countering Violent Extremism in Nigeria: Using a Text-Message Survey to Assess Radio Programs

RAND
COUNTERING VIOLENT EXTREMISM
IMPACT EVALUATION

This series examines the impact of international mass-media campaigns designed to counter violent extremism. In this report, researchers examine the effects of a campaign in northern Nigeria.

KEY FINDINGS

■ Using a randomized encouragement design, we assigned 2,064 people to listen to either *Ina Mafita* or to a control program (professional soccer matches) each week over the course of two months. Recruitment and engagement were conducted remotely via short message service (short message system [SMS] or text message).

■ The results indicate that the radio show *Ina Mafita* had a positive effect on listeners' beliefs about the importance of being a role model and a positive but not significant effect on the belief in local committees' value in reintegrating at-risk youth. Results were more pronounced for high complier subsample and for those who reported liking the show's story line.

■ The authors found no effect on listeners' views of kidnap victims.

■ The researchers found no or possibly negative effects on listeners' value of diversity, however, it must be noted that the show did not explicitly address this theme.

■ Listeners also enjoyed the show and many continued to listen to the show after the incentivized exposure had concluded.

Nigeria has long suffered from the deadly actions of Boko Haram, a locally rooted terrorist organization operating in the country's northeast region, and the Islamic State of West Africa, a local offshoot of the Islamic State of Iraq and the Levant (Blanchard, 2016). Recently, however, the country has made significant gains against these groups, dramatically reducing the militants' geographic footprint and operational capability (Council on Foreign Relations, 2019). It is now important to consolidate these gains by working to stabilize and rebuild northeast Nigeria and address numerous factors that contribute to radicalization.

Nigeria, together with the international community, is funding various efforts to address sources of instability and radicalization. One critical mechanism is the use of radio and television programs to directly reach affected communities in northern Nigeria. The effective use of such media channels can hold particular value. For example, Warren (2015) has shown that widespread access to mass media in Africa is associated with a reduction in militant violence. As Pierskalla and Hollenbach argued, "The 'soft power' of mass media enables the government to dissuade insurgent collective action through the dissemination of pro-government propaganda" (2013, p. 210).

Only a few studies, however, have examined the influence of specific radio programs that seek to *counter violent extremism* (CVE), that is, "to address the conditions and reduce the factors that most likely contribute to recruitment and radicalization by violent extremists" (U.S. Department of Homeland Security, undated). To help improve the evidence base for such radio programming, RAND Corporation researchers conducted an SMS-based randomized control encouragement study to evaluate the effect of a radio talk show intended to address underlying factors promoting instability and support for Boko Haram in northern Nigeria. This report is part of a broader RAND portfolio of evaluation studies enabled by funding from the Global Engagement Center at the U.S. Department of State.

The specific program that RAND evaluated was called *Ina Mafita* (The way forward), a talk show–style radio program produced and distributed by Equal Access International (EAI). *Ina Mafita* is one of three programs in EAI's *Farar Tattabara* (White dove) suite of Hausa-language shows, which are part of its portfolio of CVE projects launched in 2017 in northern Nigeria.[1] In 2018–2019, the program aired every week for 49 weeks in each of the 13 northern Nigerian states. The show consisted of a brief prerecorded story line that continued from week to week, followed by a live call-in session moderated by the radio host, during which listeners could share their opinions about the story line and associated themes.

Our evaluation focused on ten weeks of the show's programming in late summer and early autumn 2018. *Ina Mafita* addressed various themes using different story lines, each spanning four weeks, with a recap episode in the fifth week. Our evaluation coincided with two different story lines developed around topics for which EAI found room for improvement in the target audience's preconceived beliefs or opinions (Dietrich, Greiner, and Compton, 2018). The first story line was intended to emphasize the value of strong role models and organized local committees in orienting youth away from radicalization or in reintegrating former radicals. EAI's formative research (Dietrich, Greiner, and Compton, 2018) concluded that negative role models currently play a role in radicalizing youth, and a 2016 report recommended better connections to positive role models to counterbalance the pulls of radicalizing messages (Mercy Corps, 2016). The second story line was intended to refute the notion that kidnap victims are bad people and must have done something to deserve their abduction; rather, they are innocent and have experienced trauma and deserve sympathy and concern. Previous research has established that those who have been kidnapped or forced into marriage by Boko Haram face serious stigma when they leave the group (Felbab-Brown, 2018). Thus, efforts that can reduce the prevalence of this type of stigma could play a critical role in helping

to reunite communities that have been torn apart by Boko Haram.

Our study contributes to a growing literature on CVE program evaluation and, more broadly, on the role of messaging campaigns in promoting positive social change in areas affected by violence and conflict. We also contribute to the understanding of the specific program and region being studied, the evidence for best practices regarding mobile phone surveying, and the resulting lessons for future work.

Prior studies vary in terms of methodological rigor and ability to infer causality. Paluck and Green (2009) provide a prominent example of the use of an experimental design and an attempt to control for confounding factors, such as noncompliance. The researchers used a randomized controlled design to assess the effects of a yearlong radio soap opera on community reconciliation and intergroup prejudice in Rwanda. They found that, compared with participants in the control condition, people in the experiment group—who were systematically exposed to the treatment radio program—experienced improvements in perceptions about social norms involving intermarriage, trusting people, open dissent, and discussion of trauma. The radio program, however, did not produce substantial changes in participants' personal beliefs about intergroup prejudice, violence, and trauma. In another study, Bilali and Vollhardt (2013) conducted a field experiment in Rwanda to assess the effects of a radio drama that sought to increase openness to perspectives with regard to the history of intergroup conflict. Following the national airing of the radio drama, investigators randomly assigned participants to listen to an audio-delivered questionnaire that used either the voice of the drama's main character or an alternative voice. The investigators found that participants "primed" with the drama's voice actor were more likely to report

higher levels of historical perspective-taking and expressed less mistrust toward the outside group. The authors subsequently replicated the methodology for a show in the Democratic Republic of Congo, finding that priming had mixed effects: Respondents showed decreased social distance but also decreased confidence in the utility of intergroup dialogue (Bilali and Vollhardt, 2015).

More broadly, research suggests media programming is valuable in CVE. A recent systematic review of CVE programming found that, among the roughly eight studies that evaluate influence, most show at least some statistically significant findings. An evaluation of U.S. Agency for International Development programs was able to demonstrate success in some but not all anticipated outcomes. For example, Aldrich (2014) found that program participants in Timbuktu, Mali, were more likely than nonparticipants in neighboring areas to report listening to peaceful radio programs and participate in community decisionmaking. However, the program was not able to improve perceptions about whether the United States was fighting Islam or improve perceptions that al Qaeda activities are justified. In a Dutch CVE intervention, posttreatment questionnaires showed that three workshops designed to help migrant Muslims manage issues of identity and multiculturalism led to self-reported improvements in agency and reduced attitudes toward ideology-based violence and participants' own violent intentions. However, other key outcomes were not statistically significant (Feddes, Mann, and Doosje, 2015). Overall, these studies show that CVE initiatives hold promise and are worthy of further scientific study.

One significant critique of the literature, however, is that the available body of research is too small to effectively inform the development of new CVE programming, guide program improvements,

A recent systematic review of CVE programming found that, among the roughly eight studies that evaluate influence, most show at least some statistically significant findings.

or help policymakers and donors decide to sustain, scale up, or discontinue program activities. Rigorous evaluations are obviously critical because they help ensure that programming funds are dedicated to the most-effective efforts. Evaluations also play a critical role in helping individual programs improve the quality of service provision (Helmus et al., 2017; Helmus and Klein, 2018). Additional work in this area will be of great value in improving the effectiveness of CVE programming; to this end, we conclude this report with lessons and recommendations for future impact evaluations using randomized controlled study designs.

About This Report

The remainder of this report details our assessment methodology, findings, limitations, and recommendations. The next section describes the methodology, including survey design and implementation, empirical measurement strategy, and comparison to prior related research. The "Findings" section describes our findings, sample demographic characteristics, treatment effects, and measures of listenership. The "Discussion and Recommendations" section highlights our key takeaways and discusses limitations on the interpretation of the results. That section also lists recommendations for future work, including ways to improve the radio show design and the associated analysis.

Methodology

To assess the influence of the CVE-oriented radio program, we employed a panel study with a randomized controlled encouragement design study wholly administered via SMS technology. Participants across northern Nigeria were recruited via SMS and randomly asked to listen to ten weeks of either the treatment radio program or a control program that consisted of international soccer matches. Participants were incentivized for listening to their assigned shows, and SMS-based surveys were conducted at various intervals to assess changes in audience attitudes. This section reviews this design in detail, providing descriptions of participant recruitment, the development of the survey instrument, the design of the survey, the incentive structure, and the statistical analyses. This section begins with a brief introduction about why we selected this particular study design.

Introduction to the Study Design

To evaluate the impact of *Ina Mafita*, several aspects of the program and the geographic context needed to be addressed. First, variations of the treatment program and others like it had been airing in northern Nigeria for approximately one year. Routine listeners to the show might already have been influenced by similar themes in prior story lines or might have become regular listeners based on preexisting attitudes that made them sympathetic to the radio show's plotlines. Therefore, it might have been difficult to use cross-sectional surveys to detect improvements in attitudes in the current audience. Second, the specific reach of the radio program, or the percentage of the listening population, was and remains unknown to us. This also made the use of pre- and postexposure cross-section surveys difficult to implement. If listenership rates were too low, then

Participants across northern Nigeria were recruited via SMS and randomly asked to listen to ten weeks of either the treatment radio program or a control program that consisted of international soccer matches.

the study would lack sufficient power to demonstrate pre-post differences. If the listenership rates were too high, then it would be difficult to identify nonexposed control groups. And relatedly, we do not know how frequently people listened to the program. A large audience might have listened to the program, but if they did not listen to a sufficient number of shows, then it would be difficult to know whether exposure to the intervention was high enough to produce detectable changes in audience attitudes.

For these reasons, we implemented a remote SMS-based longitudinal panel survey with a randomized controlled encouragement design, in which participants were randomly assigned to listen to either *Ina Mafita* or a control program and were paid based on evidence of compliance with the instructions.[2] The basic idea behind an encouragement design is to obtain two similar groups of participants, separated by the largest possible difference in exposure to the content being studied, by paying the two groups to engage in different activities that deliberately do or do not expose them to the content. An encouragement design is particularly useful outside laboratory settings, when compliance is not assured. For example, in this setting, we could not force treatment group participants to listen to *Ina Mafita*, and we could not ensure that control group participants would not listen. Similar evaluation designs have been implemented in similar studies of edutainment in Africa, including for health, financial literacy, and entrepreneurship (Yoder, Hornick, and Chirwa, 1996; Paluck and Green, 2009; Bjorvatn et al., 2015; Berg and Zia, 2017).[3]

SMS surveys, like any mode of survey delivery, have advantages and disadvantages. A major advantage is that SMS can reach populations that are otherwise difficult to target, such as those in northeastern Nigeria, where insurgencies can make it difficult to access respondents in person. In addition, they are cheaper and allow for much larger samples than in-person surveys. For these reasons, mobile phone surveys (including both SMS and computer-assisted telephone interviews [CATI]) have been used in several recent studies throughout Africa (Hoogeveen, Rossi, and Sansone, 2017). The SMS design also enables more-frequent touch points with the participant population. Reminders to listen to

A major advantage is that SMS can reach populations that are otherwise difficult to target, such as those in northeastern Nigeria, where insurgencies can make it difficult to access respondents in person.

the show can be sent weekly as can quizzes on show content, which were used to test whether participants listened to their assigned program. In addition, these messages, as well as the surveys, can be delivered at precise and consistent times.

The SMS design is also minimally invasive, allowing for the intervention to be as naturalistic as possible. Participants listen to or watch the assigned program as part of their normal day, and any discussion about the show occurs naturally as a result of interactions with friends and relatives rather than as part of a focus group discussion or preorganized listening group.[4] Like in-person surveys, SMS surveys also allow for free-response questions, so qualitative information can be collected and participants' thoughts can be recorded.

Drawbacks do exist, however. SMS surveys do not allow for a back-and-forth discussion to clarify answers to free-response questions, and analyzing such responses might require some subjective decisionmaking about corrections of typos or determination of participants' intents. In addition, there is little evidence to inform important decisions for remote surveys, such as the appropriate amount of payment to participants or the appropriate length of a

survey to minimize attrition. Furthermore, researchers are only recently beginning to systematically study the effect of mobile delivery on outcomes such as sample representativeness, validity and reliability of answers, and data-entry errors.[5] The lessons that can be drawn from prior work are limited by the fact that what is true in one country or population may not hold for other regions of the world.

Nevertheless, insofar as a consensus exists regarding best practices for SMS surveys, research suggests that surveys should be short (to minimize attrition) and should consist of simple questions. It is therefore recommended that surveys avoid questions that span multiple text messages and questions that involve multiple conditions (e.g., "if yes, . . . " or "if no, . . . "). There is little evidence that the level of compensation, beyond a certain relatively low value, has any effect on respondents' participation or answers. Nonrepresentativeness is always an issue because mobile-phone ownership is not random and because reading and responding to SMS messages require a certain level of literacy.[6]

Target Audience Recruitment

We worked with Geopoll, an SMS-based survey company that maintains a panel of more than 180,000 mobile phone users in Nigeria from which it draws survey respondents and compensates them in phone credit paid directly to their mobile-phone accounts. Using demographic information on this panel, we targeted respondents ages 18–35 who reside in one of the 13 states in which *Ina Mafita* airs in northern Nigeria (Figure 1 provides a map showing participating states and comparing the recruited sample with the target population). This age group matched EAI's target audience, based on the population that has been found to be at risk of radicalization (see, for example, United Nations Development Programme, 2017).[7] Recruitment and consent took place via SMS message, with an invitation being sent to the eligible population; those who accepted the invitation received additional messages explaining the length, scope, and payment conditions of the survey.

GeoPoll recruited participants based on a first-to-respond, first-to-enroll method. Their standing panel of respondents received text invitations on a random, rolling basis; the first to consent to the study were admitted until the goal of roughly 2,000 participants was met. We found that only a handful of potential participants declined to consent, but we do not know how many received the invitation and ignored it. Our final sample consisted of 2,064 individuals who were randomized into treatment and control groups stratified by age (26 or

FIGURE 1

State of Residence for Survey Sample and Overall 18–35-Year-Old Population
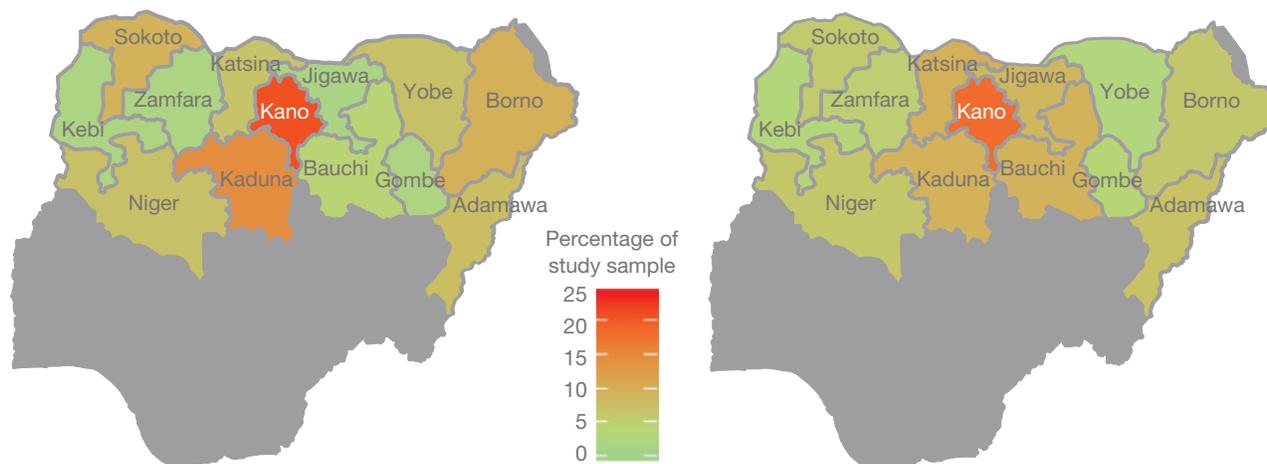
SOURCES: RAND calculations from survey data and Afrobarometer, 2016.
NOTE: Survey sample is shown on left, general population on right.

older versus 25 or younger), gender, and state of residence. This is a convenience sample drawn from GeoPoll's survey respondent panel, which is itself nonrepresentative of the Nigerian population. In the "Findings" section, we show how our sample differs from the target audience population as a whole, and in the "Discussion and Recommendations" section, we review key implications of this nonrepresentativeness.

Survey Questions and Format

The SMS format imposed some constraints on question formatting and length. The questions needed to be answerable on a flip phone without a full keyboard and without smartphone capabilities. This meant that questions needed to be short enough to fit in one SMS message, and the amount of writing required for a response needed to be minimal. Therefore, we opted for multiple-choice questions that were fewer than 160 characters in length.

The length of each survey was another constraint. As recommended by prior research and by GeoPoll, surveys should be short to maximize completion rates (Mavletova and Couper, 2015). GeoPoll recommended no more than 12 questions per wave. Some of those questions were reserved for our repeated measures of our outcomes of interest; some were used to ask for listener feedback; and some were used to collect demographic information. Because of these survey length limits, we collected time-invariant demographic information over the course of the whole study, rather than asking for all background information in the baseline wave as would be done in a face-to-face interview. Appendix A lists the survey questions and the wave in which each was asked.

Our main outcome questions addressed themes from both *Ina Mafita* story lines. Given the survey-length constraints, we were able to dedicate five questions in each wave to these outcomes. We developed the questions in consultation with EAI's formative research (Dietrich, Greiner, and Compton, 2018), their theory of change, and the scripts they developed for the show (see Appendix B for further details).[8] From these, we developed statements or questions reflecting opinions and beliefs that the show was intended to influence. Responses

consisted of five-point ordinal scales reflecting how strongly the respondent agreed or disagreed with the statement or his or her feelings about the question. The outcomes are listed in Table 1. The first two questions address themes from the first story line about the importance of role models and local committees in preventing radicalization and reintegrating radicalized youth. The second pair relates to themes from the second story line about kidnapping victims. The last relates to a broader goal of promoting tolerance and open-mindedness. It is based on a survey conducted by the Pew Research Center and provides an additional benchmark against which to ensure that our sample-response patterns are similar to those of prior studies.[9] We should note, however, that the treatment program content did not include explicit themes promoting tolerance, and any inability to detect changes on this indicator would not necessarily reflect on the effect of the treatment program.

Three additional observations on the survey-design process are important to note. First, the specific subject matter of each question was developed in an iterative manner. At our request, the program implementers provided a draft logic model detailing the connection between the content of the individual episodes and corresponding themes and goals and objectives (see Appendix B). Next, we reviewed the content for each of the ten program episodes and offered iterations on proposed objectives that could be transformed into survey questions. Once we and the implementer were in agreement on these objectives, we developed the appropriate survey questions. During the question-development and -translation process, we discussed with EAI the use of certain phrases in both English and Hausa. In particular, we verified that the concept of "role models" existed in Hausa and what phrase was used in the show. For sensitivity reasons, we avoided the use of "Boko Haram" or "violent extremists," opting instead for "rebels."[10]

Second, we then worked with translators from GeoPoll and the program implementer to translate the English questions into Hausa. This process proved difficult because neither the GeoPoll translators nor the implementer's translators could come to an agreement about the correct way to

TABLE 1
Outcome Questions Asked in Each Survey Round

Question Number	Label	Question Text
1	Role models	"It is not my responsibility to be a role model for others." • Strongly agree • Agree • Neither agree nor disagree • Disagree • Strongly disagree
2	Local committees	"Communities should form organizations to counter youth violence." • Strongly agree • Agree • Neither agree nor disagree • Disagree • Strongly disagree
3	Kidnap concern	How concerned are you for the conditions of those kidnapped by rebels? • I am extremely concerned. • I am very concerned. • I am fairly concerned. • I am not very concerned. • I am not concerned.
4	Kidnap empathy	"Women kidnapped by rebels are liable for their state." • Strongly agree • Agree • Neither agree nor disagree • Disagree • Strongly disagree
5	Diversity	"Nigeria's religious diversity is a source of strength." • Strongly agree • Agree • Neither agree nor disagree • Disagree • Strongly disagree

translate the English text into the historically oral language of Hausa.[11] We employed a translator to review and adjudicate the nonmatching translation questions, and these were reviewed with approval by the implementer's translators. As a result of the adjudication process, the final Hausa version was sometimes an approximation of the English. This was necessitated partly by character length and partly by the need for consensus among the translators. Our translators noted that question three in Hausa read closer to, "What are your feelings toward those kidnapped by rebels?" Question four in Hausa read closer to, "It is women's fault if they are kidnapped by rebels." The latter might be easier to understand than the English version while maintaining the sense of the original statement.

Third, given time constraints, we did not conduct cognitive testing on the survey questions, nor were the questions piloted in the field. These are important best practices to ensure that audiences understand the survey questions as they are intended to be understood. Although we could not go through these steps, we compared the answers for each outcome question over the five survey waves in which it was asked. For each question, Cronbach's alpha was between 0.8 and 0.9,[12] suggesting that participants' understanding of each question was consistent over the course of the survey. We are unable to say whether participants understood the questions as they were worded or whether participants interpreted the questions another way. For example, the role model and kidnap empathy questions are worded negatively, so that disagreement aligns with the show's attitudinal goals. Although anecdotal, it is reassuring that cognitive testing on similar questions for a follow-up, in-person survey in the same region suggests that there is no problem with comprehension of this type of wording in Hausa.

Survey and Incentive Structure

Figure 2 summarizes how the survey proceeded after randomization into treatment and control groups. First, a baseline survey collected demographic information and responses to each of the five outcome questions described in Table 1. Then, each group was assigned to listen to or watch a particular program for ten weeks, which was exactly the length of the two *Ina Mafita* story lines being studied. Each week, compliance was verified by asking a single quiz question relating to that week's content. Interspersed with these quizzes in weeks 5 and 10, and after exposure in weeks 15 and 32, a longer survey repeated the five outcome questions of interest, gathered more demographic information, and asked additional questions about the assigned programs. The details of each survey phase are described below.

The treatment group was told to listen to *Ina Mafita*, which aired on Sundays with particular times and stations varying by state. Control-group participants were asked to watch or listen to an English Premier League soccer match, which also aired on Sundays, always at 4:00 p.m. local time. Each Saturday, all participants received a text message reminding them of the relevant details for their assigned program in their particular location. New shows and matches were assigned every week except weeks 5 and 10 of the survey, when *Ina Mafita* aired a recap episode summarizing the prior month's story line and the Premier League had a bye week.

Ideally, control-group participants would have been assigned to listen to a different radio show that aired at the same time as *Ina Mafita* but that contained neutral messages and themes

that would not overlap with those of *Ina Mafita*. This was not possible because *Ina Mafita* aired at different times in each state and did not line up with the same programs on other radio stations. Given these circumstances, Premier League soccer games offered several benefits: They always aired on Sundays—the closest overlap with *Ina Mafita* as we could guarantee—and the bye weeks in the league overlapped with the recap weeks concluding each *Ina Mafita* story line. Most important, English soccer games would not contain any news or political messaging that might overlap with messages of *Ina Mafita*. In addition, the Premier League is popular in Nigeria, which maximized the likelihood of control group compliance—particularly among young men, who form the majority of our sample.[13]

Although the choice of control program has several benefits, some limitations should be noted. First, the control program is not precisely equivalent to the treatment: Participants might watch it on television rather than listen to the radio, a soccer game is longer than an episode of *Ina Mafita*, and participants might find the answers to the quiz questions in a newspaper or web search rather than watching the game. These differences were not critical in the context of this study, because our focus was not on control group exposure. Instead, it was important to have parallel amounts of engagement with the survey platform and to assign a task that would provide the chance to earn equivalent amounts of rewards. A more-concerning issue, raised by local EAI staff after our survey concluded, is a cultural one: Nigerians often watch soccer matches at local bars or similar establishments, and women might be excluded from these groups. Despite this concern, we

FIGURE 2

Survey Evaluation Design
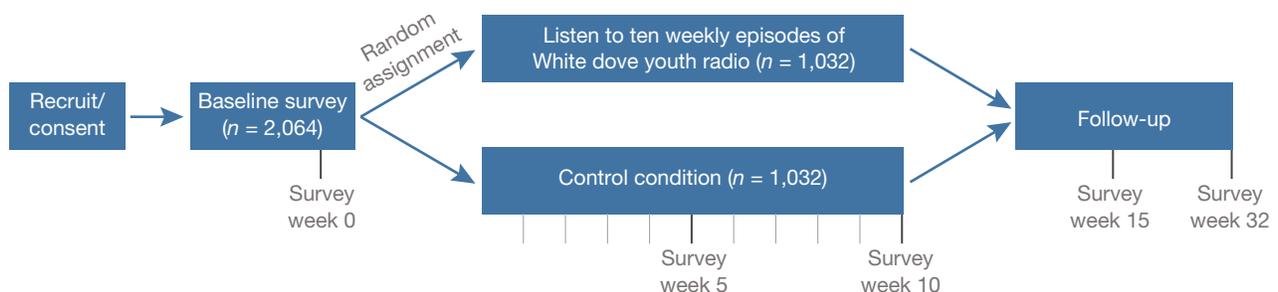
did not find differences in compliance rates between women and men in the control group.

Quizzes were sent after each new episode or soccer match (weeks 1–4 and 6–9) and consisted of a single multiple-choice question to determine whether each participant actually listened to the show or watched the game. The questions were designed to be sufficiently simple so that anyone who listened to or watched the assigned program would get them correct, but noncompliers (particularly in the treatment group) would not be able to guess or infer the answer. Appendix A lists treatment- and control-group schedules and assignments, as well as quiz questions for each group.

In weeks 5 and 10 during the exposure period, when the radio show was having its recap and the Premier League was on break, the respondents received a longer survey that repeated the five outcome questions, collected more demographic information, and gathered opinions about the assigned content. The five outcome questions were repeated again after exposure had ended, in week 15 and week 32. In those late follow-up surveys, participants were also asked about knowledge of and exposure to *Ina Mafita* prior to or since the survey to gauge cross-contamination between control and treatment groups. The timing of these longer survey waves was chosen to identify both short- and long-run effects of *Ina Mafita*'s messages: specifically, the immediate effect of the first story line before the second aired; the immediate influence of the second story line; the medium-term effect of all messages one month after exposure ended; and the longer-term impact approximately five months after exposure ended.

Compensation differed depending on the particular type of survey question. For each quiz question, participants received $1 in phone credit (approximately 360 Naira) for a correct answer. For responding to each long survey, participants were paid 50¢ (approximately 180 Naira) in phone credit. These amounts were determined in consultation with GeoPoll. Fifty cents is the typical incentive for their survey panel and was chosen to maximize participation without inducing any adverse effects, such as social-desirability bias (as in Stecklov, Weinreb, and Carletto, 2018). GeoPoll representatives also did not want to create disparities in payment for survey participation across their panels. The quiz questions were determined to require higher compensation, partly to account for the time lag between listening to the show and taking the quiz and partly to account for the relatively higher burden of listening to a 30-minute program and the associated risk of attrition over ten weeks' worth of exposure.

Comparison to Previous Studies

Our study design, exposure period, and sample size are comparable to other experimental studies of media influence, particularly those with an incentive design. Berg and Zia (2017) studied a television program's effect on financial literacy in South Africa. Although these researchers did not use SMS for recruitment or data collection, they did employ a design similar to ours, with quiz questions to gauge compliance with a sample size of about 1,000, two months of exposure, and a follow-up survey four months later. Paluck and Green (2009) studied a radio soap opera's effect on dispute resolution in Rwanda with a sample size of 40 people in each of 14 communities over the course of one year. Banerjee, La Ferrara, and Orozco (2018) studied a television show's effect on sexual behavior in Nigeria with a sample of more than 5,000 people over the course of several months. Ravallion et al. (2015) studied the influence of a film on people's knowledge of an antipoverty program in India, with more than 5,000 people in a yearlong study. Bjorvatn et al. (2015) examined the effect of an edutainment television show in Tanzania on entrepreneurship among 2,132 students with a two-year follow-up period. Bernard et al. (2015) recounted lessons from a study in Ethiopia of a documentary promoting forward-looking investment behaviors, with 2,063 participants and a six-month follow-up. DellaVigna and La Ferrara (2015) provided a more-comprehensive survey of research on media impact, including studies using other methodological approaches and focusing on other geographic regions. These studies generally show that entertainment programming can be effective in achieving the program's stated objectives.[14]

This study differs from most edutainment research in its focus on CVE narratives. As noted

in several literature reviews and policy reports, evaluations of CVE programs—media programming or otherwise—are relatively rare compared with evaluations of public health or education initiatives. Quasiexperimental evaluations are even rarer (Nasser-Eddine et al., 2011; Neumann and Kleinmann, 2013; Romaniuk, 2015; Ferguson, 2016; Khalil and Zeuthen, 2016; Mastroe and Szmania, 2016; Gielen, 2017; and Baruch et al., 2018). As several researchers have noted, this is partly because it is difficult to deduce causality in most CVE settings, rather than for lack of trying (Owens et al., 2016; Holmer, Bauman, and Aryaeinejad, 2018). The closest CVE-related model to this report appears in Aldrich (2014), who examines the effect of a suite of CVE programs in Mali, including a radio program. Also related is an experimental design by Bilali, Vollhardt, and Rarick (2017), who compared different versions of a single radio drama episode from a program in the Democratic Republic of Congo designed to promote positive change. One version of the episode models collective action, while the other does not. The emphasis on collective action resonates with the first story line of *Ina Mafita*, which promotes a community's agency in generating positive change.

Statistical Analysis

We measure our five outcome questions of interest at five key points in time: once before exposure and four times after exposure began. This allows us to estimate treatment effects over the course of exposure and, after, determine if any measurable effects persisted beyond the time that *Ina Mafita* aired. Because of the possibility of noncompliance (i.e., those assigned to the treatment group not listening to *Ina Mafita* and those assigned to the control group listening anyway), the measured effect is the *intent to treat*, which is the average difference between those assigned to the treatment and those assigned to the control.

Intent-to-Treat Effects

Our basic empirical specification is a difference-in-difference regression given by equation 1:

$$Y_{it} = \alpha + \beta T_i + \delta_t + \gamma_t(T_i \cdot \delta t) + \epsilon_{it} ,$$

where

- the constant α captures the average control group response in the baseline wave ($t = 0$)
- the dummy variable T_i records whether the individual was assigned to treatment
- the coefficient β measures the difference between treatment and control at baseline
- the dummy variable δ_t captures the change in the control group response from time 0 to time t
- γ_t is the coefficient of interest that measures the change in the treatment group from time 0 to time t that is attributable to the treatment.

The outcomes of interest are the responses to the five survey questions asking about the core *Ina Mafita* themes. Where necessary, the question responses were reverse-coded so that higher numbers align with *Ina Mafita*'s goals of greater openness, empathy, or tolerance. Because these are opinion questions with ordinal outcome scales from 1 to 5, some transformation is required to interpret the treatment effects in a consistent way across questions. We normalize the responses to each question to get random normal variables that can be compared across questions and across survey waves. Following Kling, Liebman, and Katz (2007), we calculate the mean and variance for the control-group responses to each question in the baseline wave. Then, all responses for each question are normalized using the respective mean and variance. The result is that the control group's baseline responses are standard normal variables (mean = 0, variance = 1), and the treatment effects can be interpreted as standard deviations below or above the control group's baseline.[15]

High-Complier Subgroup

In addition to analyzing intent-to-treat effects for the full sample, we looked at the set of respondents who listened to the majority of *Ina Mafita* episodes, as measured by those who answered more than 50 percent of the quiz questions correctly. There were 533 highly compliant respondents, constituting 52 percent of the treatment group. People who listened to the show often were not likely to be randomly selected from the treatment group, and

we accounted for this selection bias by matching the high compliers to observably similar control-group participants. In other words, we did not analyze the set of high compliers in the control group; instead, we identified a set of control-group participants who, as best as we can tell, would have been most likely to be high compliers had they been placed in the treatment group. This group matches the treatment high compliers along all observable characteristics.

The observationally similar control group sample was identified using propensity score matching, a standard method for constructing counterfactual samples when participation in a treatment program is nonrandom.[16] To do this, we ran a probit regression to predict high compliance in the treatment group using all available demographic information and responses to the baseline survey questions. Using the probit-regression coefficients, we predicted the probability of high compliance for both the treatment and control groups. For the latter, this is the estimated probability that they would have been frequent listeners of *Ina Mafita* had they been assigned to the treatment group. We then matched the high compliers in the treatment group to the two individuals in the control group with the closest predicted probabilities. We matched only to individuals with the same state, gender, and age group because these were the variables conditional on which the treatment was randomly assigned. To validate the match, we verified that the two matched groups do not have statistically significant differences in any observable demographics or baseline opinion responses. Table 5 shows the results of this test both before and after matching. As desired, any statistically significant differences prior to matching were eliminated in the matched sample.

Findings

This section first describes the participant sample and offers a check on the randomization and participant compliance, then the treatment effects, and concludes with a review of listener feedback on the radio program.

Sample Demographics, Randomization, and Compliance

Table 2 shows the sample demographics for the control and treatment groups, the high-compliance treatment subgroups, and (if available) the general 18–35-year-old population in the 13 states we studied. The population-level statistics are weighted means from two representative surveys, Afrobarometer (2015, round 6) and the World Values Survey (Inglehart et al., 2014, wave 6).

As expected, the survey sample is nonrepresentative of the reference population. It is skewed toward males, those in urban areas, unmarried individuals, students, and those with higher education. This is partly because these populations are more likely to have a cell phone and have sufficient literacy to participate in an SMS-based survey. Therefore, the results must be interpreted within the context of the population we are surveying and are not necessarily applicable to the wider target audience of *Ina Mafita*. Discussions with GeoPoll indicated that the high proportion of men reflects the demographics of GeoPoll's mobile phone panel. Men are also more likely to be at risk of radicalization and joining extremist groups, but our survey population differs from the profile of a typical at-risk individual in other ways: Our respondents are more educated, more urbanized, and slightly older than those who are most at risk (United Nations Development Programme, 2017). This suggests that our survey population is more likely than the general population to be sympathetic to the thematic goals of *Ina Mafita*, and we are less likely to observe a change in opinions if attitudes align with the program goals from the beginning. It also suggests that our results do not indicate the effects of *Ina Mafita* on those who are most at risk, but it might indicate the effects on those who potentially have the power to engage at-risk youth and to provide the supportive community environment promoted by the show.

The randomization of respondents into treatment and control groups appears valid. Respondents were randomized based on gender, age group, and state of residence, so there are no significant differences between groups based on these variables. For most other characteristics, the

TABLE 2

Demographic Characteristics of Control Group, Treatment Group, High-Compliance Group, and Reference Population

	Treatment Group (Percentage)	Control Group (Percentage)	p-Value for Difference	High-Compliance Treatment Group (Percentage)	All 18–35-Year-Old Hausa Speakers in the Surveyed States (Percentage)
Male	81.6%	81.4%	0.890	84.1%	47.6%[a]
Age 18–24	39.2%	38.9%	0.890	41.2%	43.8%[a]
Urban area	31.1%	36.0%	0.020	35.5%	24.1%[a]
Unemployed	24.8%	21.9%	0.120	27.0%	22.7%[a]
Married	35.1%	33.1%	0.350	33.4%	58.5%[b]
Student	54.2%	55.2%	0.640	52.5%	27.4%[b]
Listen to radio at least weekly	89.0%	89.7%	0.600	89.9%	90.3%[b]
Use internet at least weekly	72.9%	72.9%	0.990	74.5%	20.4%[a]
Use Facebook at least weekly	76.6%	74.2%	0.211	81.4%	N/A
Use Twitter at least weekly	34.4%	37.4%	0.166	40.0%	N/A
Most friends have same religion	70.6%	71.5%	0.658	73.9%	N/A
Education level					
No certificate	6.3%	5.5%	0.441	4.3%	39.3%[a]
Primary certificate	1.2%	1.3%	0.804	1.1%	25.2%[a]
Secondary certificate	31.3%	35.0%	0.087	35.2%	20.1%[a]
University/other postsecondary	51.9%	55.3%	0.140	52.6%	15.2%[a]
Postgraduate	6.4%	5.8%	0.558	6.8%	N/A
Education type					
Islamiyya	61.6%	56.0%	0.010	63.2%	N/A
Tsangaya	5.7%	5.8%	0.937	4.7%	N/A
Nonreligious	10.7%	15.3%	0.002	13.1%	N/A
Baseline outcome measures					
Role models	2.34	2.52	0.013	2.43	N/A
Local committees	4.64	4.59	0.252	4.60	N/A
Kidnap concern	4.48	4.49	0.867	4.48	N/A
Kidnap empathy	2.51	2.52	0.893	2.62	N/A
Diversity	3.69	3.60	0.170	3.60	N/A

SOURCES: RAND calculations from survey data.
NOTES: N/A = not available. Percentages do not sum to 100 because of rounding.
[a] Afrobarometer, 2015.
[b] Inglehart et al., 2014.

two groups still have no statistically significant differences at the 5-percent level. The exceptions are coming from an urban area (control group is more likely urban) and having attended a religious school (the treatment group is more likely to have attended an Islamiyya school and less likely to have attended a nonreligious school).[17] In addition, the groups have a small but significant difference in the average response to the role model question. The difference-in-difference regression approach controls for any differences in baseline opinions, measuring changes in each group relative to their respective baselines.

Figure 1 (see p. 6) shows the geographic distribution of the full survey sample (left panel) and the 18–35-year-old population as a whole (right panel). The sample is drawn primarily from central and eastern parts of northern Nigeria, with the largest numbers coming from Kano ($n = 432$) and Kaduna ($n = 304$). Relative to the general population, respondents are more likely to come from the states with the largest cities: Kano, Kaduna, Sokoto, Yobe, and Borno. It is important to note that the sample contains respondents from each state of Nigeria, including states in the far northeast that would not have been easy to access to perform a face-to-face survey because of ongoing insurgencies. There are no significant differences between treatment and control groups in terms of their geographic distribution. Relative to others in the treatment group, there are also no statistically significant differences in the distribution of high compliers.

Figure 3 shows the response rates for longer surveys and verified compliance rates (i.e., the percentage of respondents who answered the quiz correctly) for each quiz week. Overall, compliance is high, and the follow-up survey weeks showed little attrition. In the first two weeks, the sample was artificially capped at 200 respondents from each group because of a technical error; after this was corrected, compliance held steady for the treatment group at between 60 percent and 70 percent, except for week 6.[18] It is impossible to say whether the cap in the first two weeks caused some respondents to give up on completing surveys in subsequent weeks. However, in terms of the analysis, it means that, if anything, there could be more compliers in the sample than we were able to verify, and our

high-complier sample could be a subset of the true high compliers. Compliance for the control group was similar (except for a dip in week 8).[19] Response rates to the longer surveys were even higher—more than 90 percent at weeks 5, 10, and 15 for both groups. In week 32, the rate dropped to 88 percent for the treatment group and 83 percent for the control group.

Treatment Effects

Figure 4 shows the average score for each outcome question over the course of the study for both the control group (left panel) and the treatment group (right panel). The questions about local committees and kidnap concern show possible ceiling effects: The baseline was sufficiently high that the maximum possible change was not very large.

The expected timing of effects differs by question. The role model and local committees questions relate to the first story line, so we would expect effects in week 5, while weeks 10 and onward measure the persistence of these effects. The kidnap questions relate to the second story line, so we would not expect effects until week 10, with persistence being measured in weeks 15 and 32.
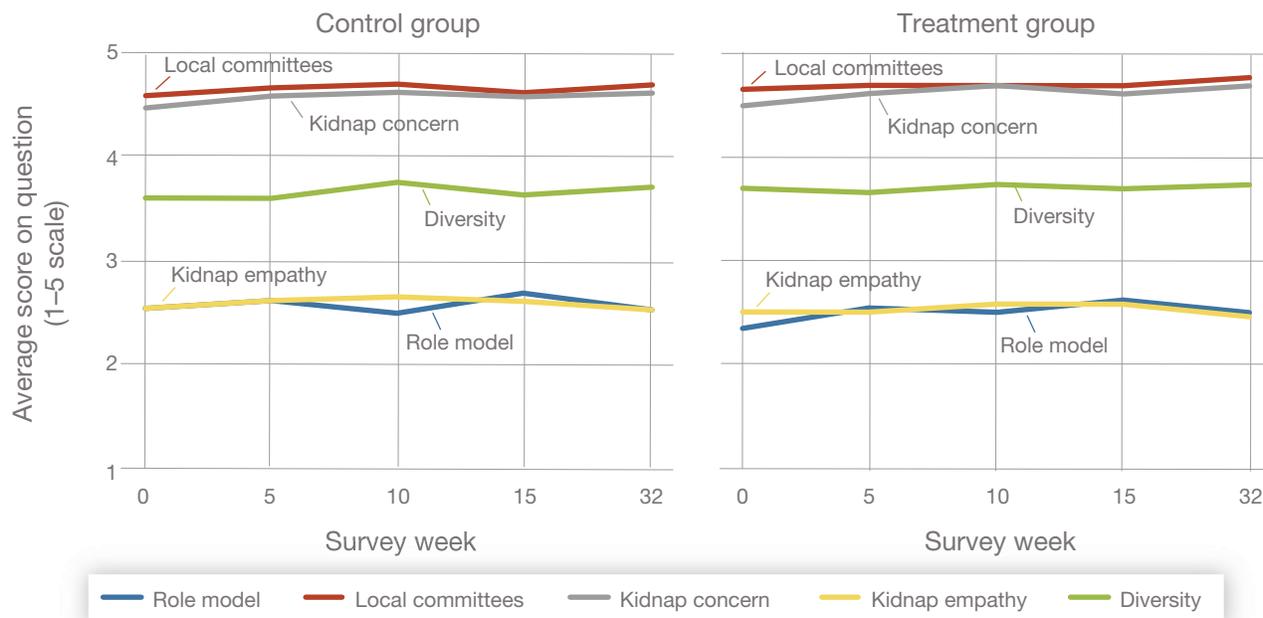
FIGURE 3

Response and Compliance Rates, by Group
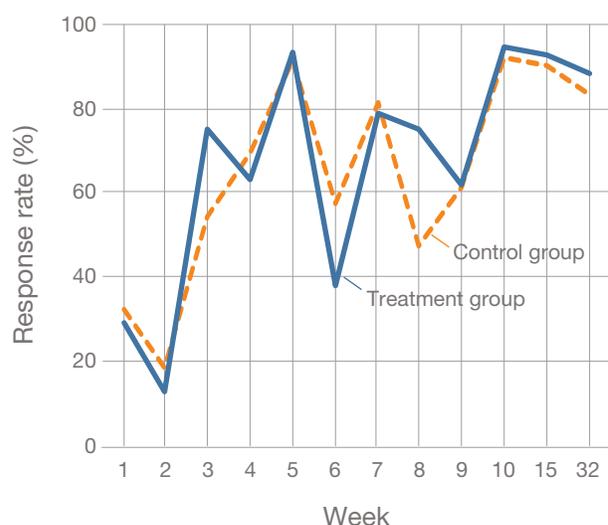
NOTE: Response rates for follow-up surveys at weeks 5, 10, 15, and 32 and verified compliance rates (percentage who answered the quiz question correctly) at weeks 1–4 and 6–9.

Average Opinion Scores for Control and Treatment Group, by Survey Wave

SOURCE: RAND calculations.
NOTES: The vertical axis shows the average response to ordinal (Likert scale) questions asking opinions of each statement in Table 1. Response scales for each question are reordered when necessary, so that 5 corresponds to agreement with *Ina Mafita*'s themes.

Figure 5 shows the treatment effects for each individual question among the full sample. Figure 6 shows the same for the match high-compliance subsample. In the full sample, the role model and kidnap concern questions show positive treatment effects, while the other three are negative; the role model treatment effect is also statistically significant at the 95-percent level in weeks 5 and 10 and weakly significant (at the 10-percent level) after (as denoted by the circular nodes and crosses). None of the other outcomes were statistically significant.

Within the matched high-compliance subsample (Figure 6), the local committees question generally has a positive instead of negative effect, although it is not significant; for the other questions, the direction is the same as for the full sample. The magnitudes of the effects are also similar. The only statistically significant effect for the high-compliance subgroup is for the role model question in week 5. The same question is weakly significant at week 32, and the diversity question has a weakly significant and negative effect in week 10.

To provide a more intuitive explanation of these treatment effects, consider the role model question,

for which the effect was approximately 0.1 standard deviations in both the full sample and the high-compliance subgroup. This is measured relative to the baseline distribution for the control group. For the role model question, the relevant standard deviation was 1.65, meaning that, for a measured effect of 0.1, the treatment group must have increased their average score by 0.165 more than the control group. Given that there are 1,032 people in each group, a change of this magnitude would require that 43 additional treatment group respondents (4.2 percent of the group) have the largest possible change, moving from a score of one to a score of five on this question. It could also occur if 170 additional respondents (16.5 percent of the group) shifted their response by just one point on the five-point scale. In other words, this effect is quite moderate, attributable to a minority of respondents having a minimal shift in opinion. In terms of standard deviations, it is, notably, comparable to effect sizes in other studies.

Overall, the treatment effects suggest that the treatment group saw a sustained improvement on two measures but not on the others. In the high-compliance group, the results suggest that the

FIGURE 5

Treatment Effect by Question, Full Sample
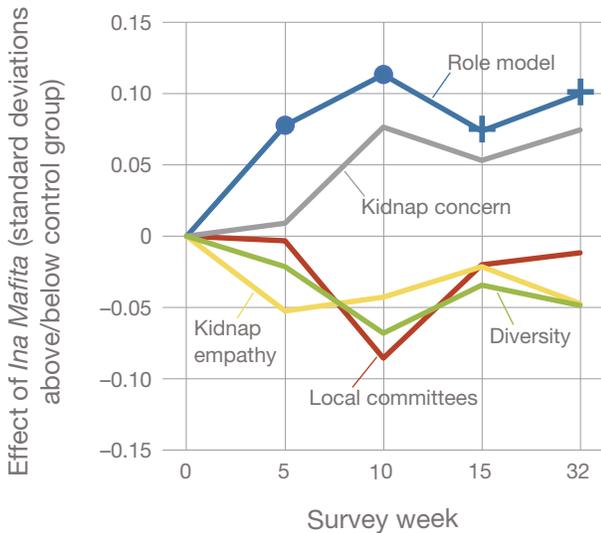
SOURCE: RAND calculations using equation 1.
NOTES: Large circular nodes indicate that the response was statistically significant at the 5-percent level for that week; crosses indicate significance at the 10-percent level.
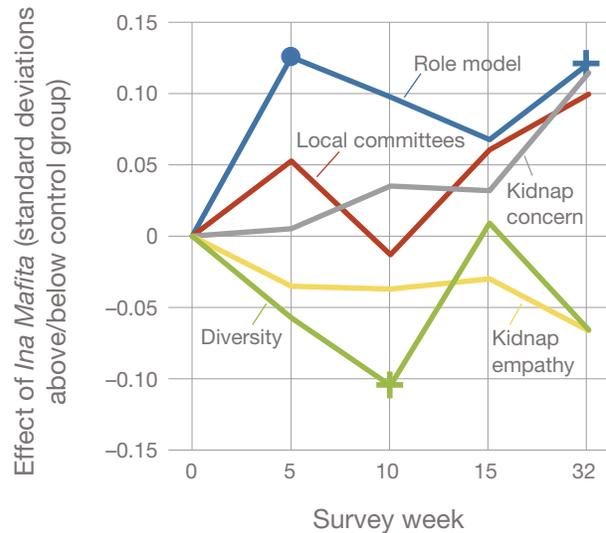
FIGURE 6

## Treatment Effects by Question, Matched High-Compliance Subsample

SOURCE: RAND calculations using equation 1.
NOTES: Large circular nodes indicate that the response was statistically significant at the 5-percent level for that week; crosses indicate significance at the 10-percent level.

first story line of *Ina Mafita* generated positive effects, but the effects of the second story line were ambiguous and possibly negative. Yet as we note in the "Discussion and Recommendations" section, it is not necessarily correct to attribute the measured effects to the radio show, and the results should be considered in the broader context of the discussion below.

Listenership and Audience Feedback

In weeks 5, 10, 15, and 32, we asked participants about their opinions of *Ina Mafita* (for weeks 5 and 10) and whether they continued listening after the incentive period concluded (week 15). The treatment group reported high rates of enjoyment, with 72 percent liking the first month and 80 percent liking the second month.[20] Among high compliers, the rates of enjoyment were 82 percent and 90 percent. As a reference point, 75 percent of the control group reported enjoying the Premier League games in the first month, and 82 percent in the second month.

It is possible that enjoyment of the show is correlated with a respondent's opinions or with the

effect of a particular story line on their responses to the survey questions. To observe whether this is true, we estimated the treatment effect for those high compliers (compared with their control group matches) who reported liking a particular story line. For the role model and local committees questions, we compared those who liked the first month with their matched controls; for the kidnap questions, we compared those who liked the second month with their matched controls. We omitted the diversity question from this calculation because it does not derive from either story line.

Figure 7 shows the resulting differences in responses. For every question, the effects are amplified compared with the full set of high compliers, sometimes twice the size of the overall treatment effects shown in Figure 7. The role model question is significant in weeks 5 and 32, and the local committees question is weakly significant at week 32. The kidnap empathy question has an even more amplified negative effect than before, weakly significant in week 32.

The results suggest that the effect of *Ina Mafita* is related to its resonance and appeal with

FIGURE 7

Treatment Effect for High Compliers Who Liked *Ina Mafita*
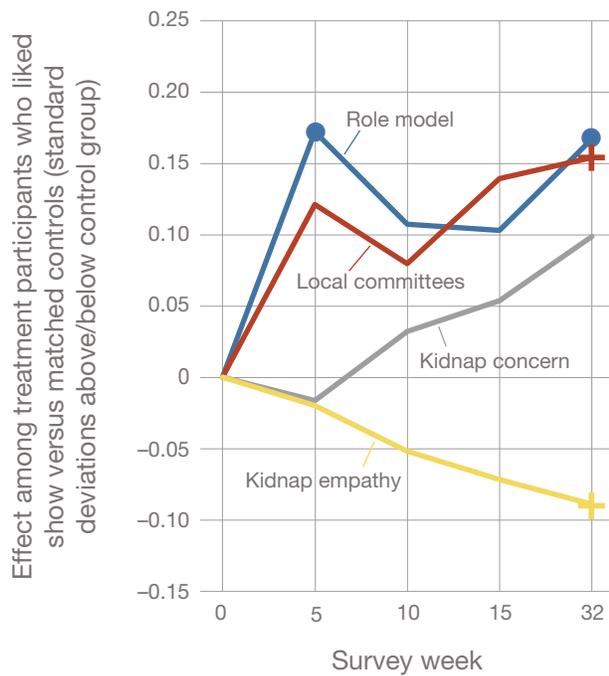
SOURCE: RAND calculations using equation 1 for high compliers who reported liking a particular month of *Ina Mafita* matched to similar control group participants.

NOTES: The role model and local committees calculations used participants who reported liking the first month of the show; the kidnap calculations used participants who liked the second month. Circular nodes indicate that the coefficient is statistically significant at the 5-percent level for that week; crosses indicate significance at the 10-percent level.

its audience, although we cannot be certain of the direction of causality. It does not appear that those who enjoyed a certain month were more likely to have particular attitudes from the beginning: We do not find statistically significant differences in baseline opinions between those who ended up liking the show and those who did not. For this subsample more than the others, the effects align with the month of the relevant story line: The effects for the first story line peak after the first month before tapering off in the second month, while the effects for the second story line do not show a shift away from zero until the second month. But overall, these trends are not measured with enough precision to say if they are statistically significant.

Insofar as enjoyment is linked to the show's influence, it is useful to understand what the audience liked. Listeners listed the top three elements

of the show that they enjoyed: 25.4 percent liked the educational and informative value (some said it was an "eye-opener"), 10.1 percent liked the story line, and 8 percent liked the main character, Zainab. Conversely, the most-listed disliked aspect, after "nothing," was the presentation, although only 28 people reported anything that they disliked.[21]

In week 15, we asked if respondents had continued listening to *Ina Mafita* or watching Premier League soccer games. Eighty-four percent of the treatment group reported listening to at least one episode since the week 10 survey (mean: 2.4 episodes); among high compliers, 89 percent watched at least one (mean: 2.7 episodes). To compare, 83 percent of the control group continued to watch soccer matches (mean: 2.4 weeks' worth of matches). These results suggest that the show was enjoyable and that people wanted to watch it, although we cannot be certain how much of the response is driven by social desirability bias. As a benchmark for the potential amount of exaggeration in responses, about 45 percent of the treatment group reported that they listened to more episodes than indicated by their quiz responses, so it is possible that a sizable fraction of the sample overreported the degree to which they continued listening to *Ina Mafita* after the incentives ended.

In the final wave at week 32, we asked both the treatment and control groups whether they had listened to any *Farar Tattabara* show (not necessarily *Ina Mafita*). Table 3 shows the overall impressions of *Farar Tattabara*. The treatment group reported a higher proportion of active listeners than the control group, which is expected. The control group also had a higher proportion of those who have never heard of *Farar Tattabara* at all. Nearly 25 percent of the control group indicated that they had listened to the show. Although it is a minority of the sample, the presence of any *Ina Mafita* listeners in the control group would attenuate the treatment effects, making our measured outcomes smaller than the "true" effects.

In addition, 26.4 percent of the treatment group still reports never having heard of the brand, despite nearly everyone in the treatment group saying they listened to at least one episode of *Ina Mafita*. These patterns suggest that the survey might have raised

TABLE 3
Self-Reported Listenership of *Farar Tattabara* at Conclusion of Survey

	Control Group	Treatment Group	High Compliers (in Treatment Group)
Never heard of *Farar Tattabara*	43.0%	26.4%	20.4%
Heard of it but never listened	32.4%	22.0%	20.0%
Listened once or twice	9.5%	14.2%	14.8%
Listened monthly	2.7%	6.3%	6.2%
Listened weekly	12.4%	31.1%	38.6%

NOTE: Calculations based on questions asking whether the respondent had heard of *Farar Tattabara* prior to this study and, if so, how often they had listened.

awareness of *Farar Tattabara* but that there is still a lack of connection in listener's minds between *Farar Tattabara* and its constituent shows.

Discussion and Recommendations

Our results indicate that *Ina Mafita* had a positive effect on listeners' beliefs about the importance of being a role model and a positive but not significant effect on the belief in local committees' value in reintegrating at-risk youth. We found zero effect on listeners' views of kidnap victims. We also found zero or possibly negative effects on listeners' value of diversity; however, it must be noted that the show did not explicitly address this theme. We found that listeners enjoyed the show, but that, in the general population, there is probably room to expand listenership by raising greater awareness of the show and the other shows under the *Farar Tattabara* brand. In this section, we discuss our results in context, including limitations of our study and its conclusions.

First, why do some questions show effects but others do not? In particular, the role model question stands out as having a sustained, positive effect over the course of the entire survey. It is possible that this question was more straightforward and comprehensible than other questions. It is also possible that this theme was addressed more directly or more consistently than the others. The talk-show format could also have influenced outcomes; for example, callers could have agreed on the importance of role models but disagreed on blaming kidnap victims.

The talk-show section of the show differed for each station, and we do not have transcripts or recordings of the show for each of the 13 states, so we cannot provide support for or against such explanations.

Second, the signs of certain effects, even if not statistically significant, also raise questions. Negative effects are difficult to explain in our causal framework and highlight the need for caution in attributing effects to a particular source. The kidnap empathy question, for example, was worded so that agreement was in opposition to the goals of the show; the same was true of the role models question, which showed a positive effect. As noted earlier, we are unable to say for certain that participants understood the questions this way, but evidence from a follow-up study in progress at the time that this study was written suggests that the wording might not have been a hindrance. The diversity question also showed negative effects, but it was not directly related to any themes from *Ina Mafita*'s story line. One conclusion would be that the show changed listener's opinions on the value of diversity, perhaps as an unintended consequence. Another conclusion could be that listeners changed their notions of what makes Nigeria strong—diversity might have taken a lower place on the list than other factors. Yet another possibility is that the show highlighted grievances and problems in listeners' communities, making them less disposed to think of outsiders or other groups as helpful in finding common ground. The latter interpretation is similar to the conclusions of Bilali, Vollhardt, and Rarick (2017), who emphasize the unintended consequences of the radio show they analyzed in the Democratic Republic of Congo. Other studies

also show that some CVE campaigns can produce iatrogenic effects among some audiences.[22]

Evaluation of Impact

The randomized control encouragement design of our survey allows for a causal interpretation of our measured treatment effects. Yet we caution against interpreting our results as the true, full impact of the show. There are several reasons for this.

First, we should note that although this study used a randomized controlled design, participants were not blinded to their study condition. Thus, participants in the treatment condition could possibly surmise the purpose of the study from connecting the content of the shows with the content of the surveys. This is a challenge that confronts all such media panel evaluations.

Second, a general confounding issue for media impact evaluations is the potential presence of other programs with similar messaging. As noted by other researchers, this is a particular problem for analyses of CVE programming, since CVE interventions are rarely implemented in isolation (Schomerus, El Taraboulsi-McCarthy, and Sandhar, 2017, p. 20). Both the treatment and control groups could be "treated" by other interventions focused on similar themes or outcomes as *Ina Mafita*, which would attenuate the difference in exposure between our two experimental groups and shrink the measured treatment effects. Moreover, the small treatment effects we measured should not be interpreted as definitive signs that the show is not doing its job: Even when the isolated effect of a single intervention (such as was the case with the *Ina Mafita*) is small, when the messaging is part of a larger, coordinated campaign, the amplification of the messaging can potentially have significant effects. Future research could assess the effect of a portfolio of CVE interventions simultaneously. In addition, when such a program is aired across a large population, even small effects can translate into still meaningful population-based change.

Next, even within our own experimental framework, there could be cross-contamination of exposure between treatment and control groups. Our follow-up survey questions did not indicate that large numbers of the control group were listening to any *Farar Tattabara* programs. Yet there could be social network effects by which the control group hears about the show or discusses the themes with friends who do listen, in which case any measured effect will be attenuated as the control group effectively becomes treated. Respondents' social networks could be studied explicitly, for example, by surveying respondents' friends and family through a snowball sampling method. CATI calls or focus groups could also uncover the importance of social-network effects.

Finally, the show might have had a larger, positive effect on opinions or beliefs that we did not measure. We designed our outcome questions to address major themes from each *Ina Mafita* story line, but, given the talk-show nature of the program, it is possible that the precise topics discussed most in-depth were different from our questions, and those topics could also have varied by state depending on callers to each radio station. As already noted, we did not have access to transcripts from each of the 13 states and thus cannot verify how exactly the show differed by state. Therefore, it is possible that, in various states, there were larger effects that remain undetected.

Lessons Learned and Recommendations for Future Work

Programming Improvements

The survey results indicate that listeners greatly enjoyed *Ina Mafita* and that they continued to listen even when not incentivized to do so. Yet among the broader population, there did seem to be room for improving listenership. This would primarily require raising awareness of the show and of the umbrella brand of *Farar Tattabara*. What is unclear from the survey is how best to do this. Given the current diffusion of technology among young Nigerians, it is possible that social-media advertising could raise interest among the target audience of young people. It is also possible that radio advertisements could raise awareness and interest among those who already listen to radio but not to *Ina Mafita*. Additional research could also provide useful information on whether the target audience would voluntarily listen to *Ina Mafita* over other programs that may air at the same time.

Assessment Improvements

Future assessments of such shows as *Ina Mafita* could be improved in several ways, both in terms of the assessment strategy and the show's design process. First, researchers should consider additional assessment strategies; the SMS survey methodology used in this study had several benefits, including low cost and access to all regions in which the show aired. The major drawbacks were the lack of a representative sample and the limitations in survey question wording and format. To obtain a more representative sample and also ask more (and more sophisticated) questions, a face-to-face survey would be desirable.[23] A CATI survey may offer some similar benefits, although any requirement that participants own mobile phones or have a baseline level of literacy may result in nonrepresentative samples. It is also possible to conduct CATI calls with participants following participation in the SMS study. GeoPoll offered to perform CATI surveys with participants in this study, but budget limitations did not allow us to take advantage of that service.

There may also be value in expanding the evaluation beyond quantitative measures. Several studies use focus groups in addition to surveys to gauge such aspects of the show as why people do or do not like the format and content.[24] CATI calls or semistructured in-person interviews can also be used to collect qualitative information. These methods can be useful in gauging listenership and helping practitioners understand the likely audience for the show. They can also help researchers identify the mechanisms through which any treatment effects work—for example, via particularly memorable messaging, forming an emotional connection with the story line, or social networks in which listeners discuss content and thereby change their opinions and attitudes through debate.

Putting the Results in Context

Overall, the findings of the study, which suggest that a CVE program can achieve some positive results, appear to be in line with the broader literature, which finds largely but not uniformly positive effects of anticonflict radio programming. Current research provides increased confidence that entertainment-focused radio programs are an effective instrument for communicating CVE or peace-focused content.

However, more research is needed to understand the true effects of such programs. A particular drawback of any CVE research agenda is the lack of a foundational literature outlining best practices or rules of thumb for designing CVE impact evaluations. It would be useful to know, for example: Are our measured treatment effects typical among similar impact-evaluation studies? Or how much exposure to the radio program is "enough?" Or which components of the radio program should producers target if they want to increase the impact? Unfortunately, there is very little research providing guidance on these points.

As noted earlier, recent literature reviews of CVE program impact evaluations conclude that rigorous, causal studies are the exception rather than the rule. This contrasts sharply with such fields as public health or development economics, in which randomized controlled trials (among other rigorous methods of causal inference) have become more commonplace.[25] There still remains a relative dearth of rigorous CVE studies, which means that the only point of comparison may be evidence from another country or another type of intervention. This is especially problematic because lessons from one region or one type of intervention may not fully apply to another; as others have pointed out, causes of violent extremism are complex and can be highly idiosyncratic (Schomerus, El Taraboulsi-McCarthy, and Sandhar, 2017).

Therefore, future CVE research would benefit from additional causal work in multiple ways: Not only would the measured effects be informative, but the lessons from the impact evaluation design—what worked, what did not, and in what context—may be equally or more valuable for other research teams.

Appendix A. Survey Instrument

Formative Research and Theory of Change

EAI conducted formative research that was documented in a report and informed the story lines that were chosen for *Ina Mafita* (Dietrich, Greiner, and Compton, 2017). Based on hundreds of interviews as well as reviews of prior research, the report identifies repeated themes in the research process and that highlight topics that CVE efforts should address. The themes touch on reasons why youth join extremist groups and on the unaddressed grievances expressed by at-risk youth and include economic and educational opportunity, levels of trust between citizens and security forces, tension between religious sects, addiction and criminality, and gender inequality. Overall, EAI's goal with the *Farar Tattabara* platform is to

> reframe radicalization to take into account human potential, unique leadership abilities, the potential predisposition of some 'radicals' towards self-efficacy, agency, and empowerment, and the need to create alternative pathways in closed societies for frustrated individuals to engage in positive social change. (Dietrich, Greiner, and Compton, 2017, p. 4)

Reorientation and reintegration of at-risk or radicalized youth is a key component of this goal and is something that EAI concludes is "a critical missing component" of current counterextremism interventions (Dietrich, Greiner, and Compton, 2017, p. 3).

EAI crafted the narratives for its various story lines to touch on the themes identified in the formative research report. EAI then provided us with a summary of the brief story line that introduces each episode of *Ina Mafita*. We asked EAI to develop a theory of change, linking the plotline from each episode to underlying knowledge, attitudinal, or behavioral goals and objectives. For example, we offered the following template:

> By helping our target audience believe (or do) _____ [short term objective],

then we can achieve a longer term goal of _____ [long-term outcome].

From the goals provided for each episode, we identified goals for each monthlong story line and developed two questions for each story line that measured attitudinal alignment with those goals.

The theme of the first month was role models. The stated objectives for the episodes were as follows:

- Listeners should understand the importance of having someone to look up to.
- Listeners should feel empowered to adopt the traits of a positive role model.
- Listeners should understand the importance of having goals and a positive outlook.
- Listeners should value the community as a major partner in reorienting and reintegrating at-risk youth.

EAI emphasized the organization of local committees at the community level as a way to identify and help reintegrate troubled youth. The role model and local committees questions (see Table 1) were developed to measure how well listeners' attitudes and beliefs aligned with these goals.

The theme of the second month was rehabilitation and reintegration, with a focus on female kidnap victims. The stated objectives for the episodes were as follows:

- Listeners should understand how women are recruited into armed opposition groups, both through force and through coercion.
- Listeners should understand how kidnapped women are treated by their abductors and why they might choose to obey certain orders even if they do not believe in the cause.
- Listeners should understand the difficulties of remaining resilient to extremist ideologies during kidnapping.
- Listeners should understand the importance of community support and destigmatization in successfully reintegrating former kidnap victims.

The kidnap concern and kidnap empathy questions (see Table 1) were developed to measure how well listeners' attitudes aligned with these goals.

Survey Questions

The surveys for weeks 0, 5, 10, 15, and 32 are reprinted here. For each wave, the survey also asked all five outcome questions listed in Table 1. At the beginning of each survey wave, respondents were asked to confirm their birth year to check that the same person was answering the questions.

Baseline (Week 0)

1. What is your gender? [For randomization]
   a. Male
   b. Female

2. What is your birth year? [For eligibility and randomization]

3. What state do you live in? [For eligibility and randomization]

4. How often do you listen to radio?
   a. Every day
   b. A few times a week
   c. A few times a month
   d. A few times a year
   e. Never

5. To what extent do you use the internet?
   a. I don't have access
   b. I have access but do not use it
   c. A few times a month
   d. A few times a week
   e. Every day
   f. Other

6. How do you usually access the internet? [Asked of those who indicated in Q2 that they use the internet]
   a. My phone
   b. Personal computer or laptop
   c. At a public building e.g., a library
   d. I use a relative/friend's internet
   e. Another method

7. Are you married?
   a. Yes
   b. No

8. What is your current employment status?
   a. Unemployed
   b. Student
   c. Employed but looking for more/different work
   d. Employed and not looking for other work

Week 5

1. How many [episodes of *Ina Mafita*/Premier League matches] did you [listen to/watch] this past month?

2. How much did you like [*Ina Mafita*/Premier League]?
   a. I liked it very much
   b. I liked it a little bit
   c. No opinion
   d. I did not like it
   e. I greatly disliked it

3. What did you like most about [*Ina Mafita*/Premier League]? [Free response]

4. What did you like least about [*Ina Mafita*/Premier League]? [Free response]

5. Would you describe your neighborhood as rural or urban?
   a. Mostly rural
   b. Mostly urban
   c. A mix of rural and urban areas

Week 10

1. How many [episodes of *Ina Mafita*/Premier League matches] did you [listen to/watch] this past month?

2. How much did you like [*Ina Mafita*/Premier League]?
   a. I liked it very much
   b. I liked it a little bit
   c. No opinion
   d. I did not like it
   e. I greatly disliked it

3. What did you like most about [*Ina Mafita*/Premier League]? [Free response]

4. What did you like least about [*Ina Mafita*/Premier League]? [Free response]

5. What is your highest level of education?
   a. No degree/certificate
   b. Primary certificate
   c. Secondary certificate
   d. Degree
   e. Advanced degree

6. Did you go to an Islamiyya or Tsangaya school?
   a. I went to Islamiyya.
   b. I went to Tsangaya.
   c. I went to a different religious school.
   d. I didn't go to a religious school.

Week 15

1. How many [episodes of *Ina Mafita*/weeks of Premier League matchups] did you [listen to/watch] in the past month?
2. How often do you use Facebook?
   a. I don't have a Facebook profile.
   b. A few times a year
   c. A few times a month
   d. A few times a week
   e. I use it every day.
3. How often do you use Twitter?
   a. I don't have a Twitter account.
   b. A few times a year
   c. A few times a month
   d. A few times a week
   e. I use it every day.
4. What types of music do you listen to?
   a. Western/pop
   b. Local traditional music
   c. Both Western/pop and traditional
   d. Neither
5. What fraction of your friends are of the same religion as you?
   a. None
   b. Some but less than half
   c. More than half but not all
   d. All of them

Week 32

1. Do you think violence is justified in advancing your religion's aims?
   a. Frequently justified
   b. Sometimes justified
   c. Rarely justified
   b. Never justified
2. Have you aided extremists' victims in the past month (e.g. financially or emotionally)?
   a. Not at all
   b. Yes, on 1 or 2 days
   c. Yes, on 3 or 4 days
   d. Yes, on 5 or more days
3. Prior to this study, had you ever listened to any shows by *Farar Tattabara*?
   a. I have never heard of it.
   b. I have heard of it but never listened.
   c. I have listened.
4. How often did you listen to a *Farar Tattabara* radio show?
   a. I listened once or twice.
   b. I listened monthly.
   c. I listened weekly.

Quiz Questions

Table 4 lists the questions that were asked each week to verify compliance. Correct answers are printed in bold. Participants who answered the question correctly received $1 in mobile phone credit.

TABLE 4
Quiz Questions

Week (Date)	Assigned Premier League Game	Control Group Question	Treatment Group Question
1 August 12	Arsenal v. Manchester City	Which team scored the first goal in the game? Arsenal **Manchester City** Neither team scored	What is the name of Zainab's male friend? Salimah **Zakkariya** Zaytun Ibrahim
2 August 19	Brighton & Hove Albion v. Manchester United	How many total goals were made by Manchester United? [Free response, **correct answer: 2**]	What job does Zainab get? Teacher Housekeeper **Seamstress** Nanny
3 August 26	Newcastle United v. Chelsea	How many total yellow cards were given out in the game? [Free response, **correct answer: 4**]	Why does Zakariyya say he avoided Zainab? He was in prison. He moved away. He was sick. **He was ashamed of joining a gang.**
4 September 2	Watford v. Tottenham Hotspur	Which team was in the lead at halftime? Watford Tottenham Hotspur **The teams were tied at halftime**	What did Zakariyya do as part the gang? **Keep weapons at his house.** Kill innocent people. Injure people. Pay money to the gang.
6 September 16	Everton v. West Ham United	Which team scored the first goal in the game? Everton **West Ham United** Neither team scored	What is the name of Hafsat's husband, who was killed? Hasan Jamil Rafiq **Idi**
7 September 23	Arsenal v. Everton	How many total goals were made by Arsenal? [Free response, **correct answer: 2**]	Why does a man give Hafsat money? It is payment for housework. It is a bribe. **It is a dowry.** It is a gift.
8 September 30	Cardiff City v. Burnley	Which team was in the lead at halftime? Cardiff City Burnley **They were tied at halftime.**	An older woman gives Hafsat advice. How many years has that woman lived with her kidnappers? [Free response; **correct answer: 5**]
9 October 6	Liverpool v. Manchester	How many total yellow cards were given out in the game? [Free response, **correct answer: 4**]	Who helps Hafsat escape from her kidnappers? The old woman The police **Her childhood friend** Her former neighbor

NOTE: Correct quiz answers are in bold. The survey was administered in 2018.

Appendix B. High Complier Propensity Score Matching

Table 5 shows the results of a balance test for the matched high-complier sample. For each observable variable used in the matching algorithm, the table shows the average for the high compliers in the treatment group, the average of the matched set of control group participants, and the *p*-value for a test of the difference in means. None of the differences is statistically significant at the 5-percent level, indicating that the matched treatment and control samples are similar in every observable characteristic.

TABLE 5

Balance Test for High Compliers in Treatment Group and Matched Controls

	Mean		
Variable	High Compliers (Treatment)	Matched Controls	*p*-Value for Difference Between Means
Propensity score	0.591	0.593	0.738
Q1: Role model	2.430	2.540	0.263
Q2: Local committees	4.600	4.570	0.607
Q3: Kidnap concern	4.480	4.430	0.408
Q4: Kidnap empathy	2.620	2.600	0.872
Q5: Diversity	3.600	3.510	0.325
Mostly urban area	0.356	0.388	0.269
Mostly rural area	0.207	0.191	0.521
Prefer Western/pop	0.068	0.088	0.243
Prefer traditional music	0.215	0.182	0.185
Listen to neither	0.057	0.071	0.356
Most but not all friends have same religion	0.620	0.604	0.609
Some but less than half of friends have same religion	0.044	0.042	0.859
No friends have same religion	0.211	0.227	0.523
No Facebook profile	0.036	0.025	0.282
Use Facebook few times a year	0.049	0.053	0.780
Use Facebook few times a month	0.068	0.063	0.725
Use Facebook few times a week	0.215	0.201	0.587
Use Facebook every day	0.601	0.629	0.447
No Twitter account	0.283	0.289	0.838

TABLE 5—CONTINUED

	Mean		
Variable	High Compliers (Treatment)	Matched Controls	*p*-Value for Difference Between Means
Use Twitter few times a year	0.143	0.137	0.790
Use Twitter few times a month	0.146	0.138	0.691
Use Twitter few times a week	0.202	0.184	0.482
Use Twitter every day	0202	0.223	0.387
No internet access	0.103	0.110	0.702
Don't use internet	0.048	0.048	0.962
Use internet a few times a month	0.021	0.018	0.738
Use internet a few times a week	0.131	0.096	0.077
Use internet every day	0.616	0.660	0.139
Internet on phone	0.719	0.708	0.700
Internet on personal computer	0.070	0.064	0.682
Internet on friend's computer	0.021	0.027	0.523
Internet at public building	0.019	0.025	0.506
Internet via other means	0.089	0.067	0.168
Don't listen to radio	0.010	0.005	0.360
Listen to radio a few times a year	0.029	0.028	0.951
Listen to radio a few times a month	0.064	0.061	0.833
Listen to radio a few times a week	0.283	0.331	0.095
Listen to radio every day	0.614	0.573	0.177
Married	0.333	0.297	0.214
Fully employed	0.036	0.029	0.486
Employed but looking for more	0.167	0.197	0.214
Unemployed	0.272	0.236	0.179
Student	0.525	0.539	0.647
Primary certificate	0.011	0.009	0.642
Secondary certificate	0.351	0.320	0.282
University degree	0.527	0.566	0.205
Advanced degree	0.067	0.055	0.453
Islamiyya	0.635	0.632	0.932
Tsangaya	0.048	0.051	0.776
Nonreligious school	0.133	0.135	0.928

NOTE: This table shows means for high compliers in the treatment group and the controls that were matched to the high compliers based on propensity score. Measurements for opinion questions Q1 through Q5 were from the baseline survey (week 0).

Notes

[1] The other two shows are *Ilimi Abin Nema* (Pursuit of knowledge), about religious education, and *Labarin Aisha* (Aisha's tale), a radio drama dealing with various themes.

[2] The randomization is to the treatment group versus the control group. The payment for compliance is not randomized, but rather is the same for every participant, so long as they successfully verify compliance.

[3] In some studies, the incentive design is called an *encouragement design*.

[4] Listening groups or organized screenings are common in this literature (see Paluck and Green, 2009; and Banerjee, La Ferrara, and Orozco-Olvera, 2019). In contrast, Berg and Zia (2017) conduct a naturalistic intervention like ours, although they use in-person follow-up surveys.

[5] On the question of validity and reliability, see Ballivian, Azevedo, and Durbin (2015), which presents evidence from Peru and Honduras. On data-entry accuracy, see evidence from India in Patnaik, Brunskill, and Thies (2009).

[6] In addition to studies already cited, see Leidich et al. (2019) for evidence from Kenya; Hoogeveen et al. (2014) for evidence from Tanzania; and Link et al. (2014) on surveys using mobile technology. Mavletova and Couper (2015) provides a meta-analysis of 14 web surveys (not necessarily SMS). Dabalen et al. (2016) provides more-comprehensive guidance and literature reviews for mobile surveys in general, including SMS.

[7] EAI's target audience also includes youth under 18 years of age, but because of limitations on consent of minors, we did not include younger individuals in our sample.

[8] For more on theory of change, see Stein and Valters (2012).

[9] Pew Research Center, 2017. Pew's question was worded as follows: "Overall, do you think having people of many different backgrounds, such as different ethnic groups, religions, and races, makes our country a better place to live or a worse place to live?"

[10] We recognize, however, that some participants may not have made the connection between the term *rebel* and Boko Haram militants and violent extremists.

[11] One factor that may have contributed to these translation issues is that the Hausa language has regional variants and dialects that include "Kano Hausa," "Western Hausa," "Northern Hausa," "Southern Hausa," "Eastern Hausa," "Ghanian Hausa," and "non-native Hausa." These differences may have influenced the competing translations we received (see University of California, Los Angeles, Teaching and Research African Languages, undated).

[12] Cronbach's alpha is a measure used to assess the reliability, or internal consistency, of a set of scale or test items.

[13] A 2017 survey of 1,000 Nigerians found that the majority follow soccer and, of those who do, 78 percent follow the English Premier League, while 47 percent reported following the Nigerian Professional Football League (NOI Polls, 2017). Another survey found that 80 percent of those surveyed watch the English Premier League more than any other, and 12 percent said the same of the Nigerian Professional Football League (Onwumechili and Oloruntola, 2014).

[14] See also Moyer-Gusé, 2008.

[15] Treating the outcomes as continuous variables does not fully alleviate concerns regarding ordinal outcome measures, particularly the question of whether the underlying distance between consecutive categories is always the same. To address such concerns, we also replicated the analysis using latent factor methods from item-response theory (see Samejima, 2019, and, for a recent, related example, see Frounfelker et al., 2019) that account for possibly unequally spaced categories. The results were nearly identical to those presented here and therefore are omitted from the report.

[16] A foundational study is Rosenbaum and Rubin (1985). Aldrich (2014) uses propensity score matching in a similar way to match treatment participants with similar control participants. For an overview of propensity score matching as it is used across the social sciences, see Thoemmes and Kim, 2011.

[17] We ask about attendance at either of two types of religious schools. Islamiyya schools incorporate elements of Western education and traditional Islamic education. Tsangaya schools, usually boarding schools, have curricula focused exclusively on Quranic study and Islamic values.

[18] The overall response rate to quizzes was very close to the correct answer rate: Upward of 95 percent of respondents to each quiz got the quiz question correct. This suggests that people were only responding if they knew the answer and that very few people who were unsure bothered to hazard a guess.

[19] We are unable to explain why compliance dipped at certain points, given that the timing of the dips differs for treatment and control groups and that, to the best of our knowledge, they did not occur during religious or national holidays.

[20] In the first month, 25 percent of participants did not answer the question at all, 1 percent reported having no opinion, and 2 percent disliked it. In the second month, 18 percent did not respond, less than 1 percent had no opinion, and less than 1 percent disliked *Ina Mafita*.

[21] These are categorizations of respondents' answers to open-ended response questions that asked what they liked and disliked most about the show.

[22] For a review, see Darden, 2018.

[23] There are drawbacks to face-to-face surveys as well. Among other issues, the quality of the data could depend on the quality of the enumerators, and interviewer bias can distort response patterns. Face-to-face surveys are also more expensive than remote survey methods.

[24] For instance, Berg and Zia (2017) use focus groups to provide evidence that emotional connections to the characters helped to explain why certain messages from their show were more memorable than others.

[25] See for example, Berg and Zia (2017), Banerjee, La Ferrara, and Orozco-Olvera (2019), Ravallion et al. (2015), Bjorvatn et al. (2015), and DellaVigna and La Ferrara (2015). For a historical overview of the rise of randomized controlled trials in development economics, see Banerjee, Duflo, and Kremer (2016).

References

Afrobarometer, homepage, Nigeria, round 6, 2015. As of September 16, 2019:
http://www.afrobarometer.org/data/nigeria-round-6-data-2015

Aldrich, Daniel P., "First Steps Towards Hearts and Minds? USAID's Countering Violent Extremism Policies in Africa," *Terrorism and Political Violence*, Vol. 26, No. 3, 2014, pp. 523–546.

Ballivian, Ampara, João Pedro Azevedo, and Will Durbin, "Using Mobile Phones for High-Frequency Data Collection," in Daniele Toninelli, Robert Pinter, and Pablo de Pedraza, eds., *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies*, London: Ubiquity Press, 2015, pp. 21–39.

Banerjee, Abhijit, Esther Duflo, and Michael Kremer, "The Influence of Randomized Controlled Trials on Development Economics Research and on Development Policy," paper prepared for *The State of Economics, the State of the World* proceedings volume, September 11, 2016. As of September 17, 2019:
https://economics.mit.edu/files/16473

Banerjee, Abhijit, Eliana La Ferrara, and Victor H. Orozco-Olvera, "The Entertaining Way to Behavioral Change: Fighting HIV with MTV," Cambridge, Mass.: National Bureau of Economic Research, Working Paper 26096, July 2019. As of September 17, 2019:
https://www.nber.org/papers/w26096.pdf

Baruch, Ben, Tom Ling, Rich Warnes, and Joanna Hofman, "Evaluation in an Emerging Field: Developing a Measurement Framework for the Field of Counter-Violent-Extremism," *Evaluation*, Vol. 24, No. 4, 2018, pp. 475–495.

Berg, Gunhild, and Bilal Zia, "Harnessing Emotional Connections to Improve Financial Decisions: Evaluating the Impact of Financial Education in Mainstream Media," *Journal of the European Economic Association*, Vol. 15, No. 5, October 2017, pp. 1025–1055.

Bernard, Tanguy, Stefan Dercon, Kate Orkin, and Alemayehu Seyoum Taffesse, "Will Video Kill the Radio Star? Assessing the Potential of Targeted Exposure to Role Models Through Video," *World Bank Economic Review*, Vol. 29, Supplement 1, 2015, pp. S226–S237.

Bilali, Rezarta, and Johanna Ray Vollhardt, "Priming Effects of a Reconciliation Radio Drama on Historical Perspective-Taking in the Aftermath of Mass Violence in Rwanda," *Journal of Experimental Social Psychology*, Vol. 49, No. 1, January 2013, pp. 144–151.

Bilali, Rezarta, and Johanna Ray Vollhardt, "Do Mass Media Interventions Effectively Promote Peace in Contexts of Ongoing Violence? Evidence from Eastern Democratic Republic of Congo," *Peace and Conflict: Journal of Peace Psychology*, Vol. 21, No. 4, November 2015, p. 604.

Bilali, Rezarta, Johanna Ray Vollhardt, and Jason Ray David Rarick, "Modeling Collective Action Through Media to Promote Social Change and Positive Intergroup Relations in Violent Conflicts," *Journal of Experimental Social Psychology*, Vol. 68, January 2017, pp. 200–211.

Bjorvatn, Kjetil, Alexander W. Cappelen, Linda Helgesson Sekei, Erik Ø. Sørensen, and Bertil Tungodden, "Teaching Through Television: Experimental Evidence on Entrepreneurship Education in Tanzania," Bergen, Norway: Norwegian School of Economics Discussion Paper, March 2015.

Blanchard, Lauren Ploch, *Nigeria's Boko Haram: Frequently Asked Questions*, Washington, D.C.: Congressional Research Service, R43558, March 29, 2016.

Council on Foreign Relations, "Global Conflict Tracker," last updated January 14, 2020. As of January 23, 2020:
https://www.cfr.org/interactive/global-conflict-tracker/conflict/boko-haram-nigeria

Dabalen, Andrew, Alvin Etang, Johannes Hoogeveen, Elvis Mushi, Youdi Schipper, and Johannes von Engelhardt, *Mobile Phone Panel Surveys in Developing Countries: A Practical Guide for Microdata Collection*, Washington, D.C.: World Bank Group, 2016.

Darden, Jessica Trisko, "Compounding Violent Extremism? When Efforts to Prevent Violence Backfire," *War on the Rocks*, June 6, 2018.

DellaVigna, Stefano, and Eliana La Ferrara, "Economic and Social Impacts of the Media," in Simon Anderson, Joel Waldfogel, and David Strömberg, eds., *Handbook of Media Economics*, Vol. 1, Oxford, UK: North-Holland (Elsevier), 2015, pp. 723–768.

Dietrich, Kyle, Karen Greiner, and Samuel Compton, *Assessing Conflict Drivers and Reframing Radicalization in Northern Nigeria: Formative Research Report for the Farar Tattabara ("White Dove") Radio Project*, San Francisco, Calif.: Equal Access International, November 6, 2018.

Feddes, Allard R., Liesbeth Mann, and Bertjan Doosje, "Increasing Self-Esteem and Empathy to Prevent Violent Radicalization: A Longitudinal Quantitative Evaluation of a Resilience Training Focused on Adolescents with a Dual Identity," *Journal of Applied Social Psychology*, Vol. 45, No. 7, 2015, pp. 400–411.

Felbab-Brown, Vanda, "In Nigeria, We Don't Want Them Back: Amnesty, Defectors' Programs, Leniency Measures, Informal Reconciliation, and Punitive Responses to Boko Haram," in *The Limits of Punishment: Transitional Justice and Violent Extremism*, Tokyo: United Nations University, May 2018. As of September 17, 2019:
https://www.brookings.edu/research/in-nigeria-we-dont-want-them-back/

Ferguson, Kate, *Countering Violent Extremism Through Media and Communication Strategies: A Review of the Evidence*, Cambridge, UK: Partnership for Conflict, Crime, and Security Research, March 1, 2016.

Frounfelker, Rochelle L., Thomas Frissen, Isabella Vanorio, Cecile Rousseau, and Leen d'Haenens, "Exploring the Discrimination-Radicalization Nexus: Empirical Evidence from Youth and Young Adults in Belgium," *International Journal of Public Health*, Vol. 64, No. 6, 2019, pp. 897–908.

Gielen, Amy-Jane, "Countering Violent Extremism: A Realist View for Assessing What Works, for Whom, in What Circumstances, and How?" *Terrorism and Political Violence*, 2017, pp. 1–19.

Helmus, Todd C., and Kurt Klein, *Assessing Outcomes of Online Campaigns Countering Violent Extremism: A Case Study of the Redirect Method,* Santa Monica, Calif.: RAND Corporation, RR-2813-GNF, 2018. As of September 17, 2019:
https://www.rand.org/pubs/research_reports/RR2813.html

Helmus, Todd C., Miriam Matthews, Rajeev Ramchand, Sina Beaghley, David Stebbins, Amanda Kadlec, Michael A. Brown, Aaron Kofner, and Joie D. Acosta, *RAND Program Evaluation Toolkit for Countering Violent Extremism*, Santa Monica, Calif.: RAND Corporation, TL-243-DHS, 2017. As of September 17, 2019:
https://www.rand.org/pubs/tools/TL243.html

Holmer, Georgia, Peter Bauman, and Kateira Aryaeinejad, *Measuring Up: Evaluating the Impact of P/CVE Programs*, Washington, D.C.: United States Institute of Peace, 2018.

Hoogeveen, Johannes, Kevin Croke, Andrew Dabalen, Gabriel Demombynes, and Marcelo Giugale, "Collecting High Frequency Panel Data in Africa Using Mobile Phone Interviews," *Canadian Journal of Development Studies/Revue Canadienne d'études du développement*, Vol. 35, No. 1, 2014, pp. 186–207.

Hoogeveen, Johannes G., Mariacristina Rossi, and Dario Sansone, "Leaving, Staying, or Coming Back? Migration Decisions During the Northern Mali Conflict," Washington, D.C.: World Bank Group, Poverty and Equity Global Practice Group, Policy Research Working Paper 8012, March 2017.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, and B. Puranen et al., eds., "World Values Survey: Round Six–Country-Pooled Datafile Version," Madrid: JD Systems Institute, September 12, 2018. As of September 17, 2019:
http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp

Khalil, James, and Martine Zeuthen, *Countering Violent Extremism and Risk Reduction: A Guide to Programme Design and Evaluation*, London: Royal United Services Institute, Whitehall Report 2-16, June 2016.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, Vol. 75, No. 1, January 2007, pp. 83–119.

Leidich, Aimee, Samuel Kamande, Khadija Hassanali, Nicholas Owsley, David Clarance, and Chaning Jang, "How to Improve SMS Survey Response Quality and Completion Rates in Emerging Markets," Medium.com, January 16, 2019. As of June 27, 2019:
https://medium.com/msurvey-pulse/how-to-improve-sms-survey-response-quality-and-completion-rates-in-emerging-markets-cb9c8775baae

Link, Michael W., Joe Murphy, Michael F. Schober, Trent D. Buskirk, Jennifer Hunter Childs, and Casey Langer Tesfaye, "Mobile Technologies for Conducting, Augmenting and Potentially Replacing Surveys: Executive Summary of the AAPOR Task Force on Emerging Technologies in Public Opinion Research," *Public Opinion Quarterly*, Vol. 78, No. 4, Winter 2014, pp. 779–787.

Mastroe, Caitlin, and Susan Szmania, *Surveying CVE Metrics in Prevention, Disengagement and Deradicalization Programs: Report to the Office of University Programs, Science and Technology Directorate, Department of Homeland Security*, College Park, Md.: National Consortium for the Study of Terrorism and Responses to Terrorism (START), March 2016.

Mavletova, Aigul, and Mick P. Couper, "A Meta-Analysis of Breakoff Rates in Mobile Web Surveys," in Daniele Toninelli, Robert Pinter, and Pablo de Pedraza, eds., *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies*, London: Ubiquity Press, 2015, pp. 81–98.

Mercy Corps, *Motivations and Empty Promises: Voices of Former Boko Haram Combatants and Nigerian Youth*, Portland, Oreg., and Edinburgh, Scotland: Mercy Corps, April 2016.

Moyer-Gusé, Emily, "Toward a Theory of Entertainment Persuasion: Explaining the Persuasive Effects of Entertainment-Education Messages," *Communication Theory*, Vol. 18, No. 3, August 2008, pp. 407–425.

Nasser-Eddine, Minerva, Bridget Garnham, Katerina Agostino, and Gilbert Caluya, *Countering Violent Extremism (CVE) Literature Review*, Edinburgh, South Australia: Counter Terrorism and Security Technology Centre, Defence Science and Technology Organisation, Report No. DSTO-TR-2522, 2011.

Neumann, Peter, and Scott Kleinmann, "How Rigorous Is Radicalization Research?" *Democracy and Security*, Vol. 9, No. 4, 2013, pp. 360–382.

NOI Polls, "Nigerians Favour Manchester United FC to Win 2017/2018 English Premier League," September 19, 2017. As of June 24, 2019:
https://noi-polls.com/nigerians-favour-manchester-united-fc-to-win-2017-2018-english-premier-league/

Onwumechili, Chuka, and Sunday Oloruntola, "Transnational Communications, Attitudes and Fan Identity: Studying Nigeria Post-Media Reform," *Soccer and Society*, Vol. 15, No. 3, 2014, pp. 389–410.

Owens, Ross, Jonathan Evans, Jennifer Foley, and Ji Sun Lee, *Countering Violent Extremism—Developing a Research Roadmap: Literature Review*, prepared for the U.S. Department of Homeland Security, Science and Technology Directorate, Research Triangle Park, N.C.: RTI International, RTI Project Number 0214428.003.000, April 2016.

Paluck, Elizabeth Levy, and Donald P. Green, "Deference, Dissent, and Dispute Resolution: An Experimental Intervention Using Mass Media to Change Norms and Behavior in Rwanda," *American Political Science Review*, Vol. 103, No. 4, November 2009, pp. 622–644.

Patnaik, Somani, Emma Brunskill, and William Thies, "Evaluating the Accuracy of Data Collection on Mobile Phones: A Study of Forms, SMS, and Voice," in *2009 International Conference on Information and Communication Technologies in Development (ICTD) Proceedings*, Piscataway, N.J.: IEEE Press, 2009.

Pew Research Center, "Spring 2017 Global Attitudes Survey," question 43, Washington, D.C., 2017.

Pierskalla, Jan H., and Florian M. Hollenbach, "Technology and Collective Action: The Effect of Cell Phone Coverage on Political Violence in Africa," *American Political Science Review*, Vol. 107, No. 2, May 2013, pp. 207–224.

Ravallion, Martin, Dominique van de Walle, Puja Dutta, and Rinku Murgai, "Empowering Poor People Through Public Information? Lessons from a Movie in Rural India," *Journal of Public Economics*, Vol. 132, December 2015, pp. 13–22.

Romaniuk, Peter, *Does CVE Work? Lessons Learned from the Global Effort to Counter Violent Extremism*, Goshen, Ind.: Global Center on Cooperative Security, September 2015.

Rosenbaum, Paul R., and Donald B. Rubin, "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score," *American Statistician*, Vol. 39, No. 1, February 1985, pp. 33–38.

Samejima, Fumiko, "Graded Response Models," in Wim J. van der Linden, ed., *Handbook of Item Response Theory*, Vol. I: *Models*, New York: CRC Press, February 7, 2019, pp. 95–108.

Schomerus, Mareike, Sherine El Taraboulsi-McCarthy, with Jassi Sandhar, "Countering Violent Extremism (Topic Guide)," Birmingham, UK: GSDRC, University of Birmingham, 2017.

Stecklov, Guy, Alexander Weinreb, and Calgero Carletto, "Can Incentives Improve Survey Data Quality in Developing Countries? Results from a Field Experiemnt in India," *Royal Statistical Society*, Vol. 181, No. 4, October 2018, pp. 1033–1056.

Stein, Danielle, and Craig Valters, *Understanding 'Theory of Change' in International Development: A Review of Existing Knowledge*, London: Justice and Security Research Programme and the Asian Foundation, JSRP Paper 1, August 2012.

Thoemmes, Felix J., and Eun Sook Kim, "A Systematic Review of Propensity Score Methods in the Social Sciences," *Multivariate Behavioral Research*, Vol. 46, No. 1, 2011, pp. 90–118.

United Nations Development Programme, *Journey to Extremism in Africa: Drivers, Incentives, and the Tipping Point for Recruitment*, 2017.

University of California, Los Angeles, Teaching and Research: African Languages, "Hausa Language Variation and Dialects," webpage, undated. As of December 17, 2019: http://aflang.humanities.ucla.edu/language-materials/chadic-languages/hausa/hausa-language-variation-dialects/

U.S. Department of Homeland Security, Countering Violent Extremism Task Force, "What Is CVE?" webpage, undated. As of September 3, 2019: https://www.dhs.gov/cve/what-is-cve

Warren, T Camber, "Explosive Connections? Mass Media, Social Media, and the Geography of Collective Violence in African States," *Journal of Peace Research*, Vol. 52, No. 3, 2015, pp. 297–311.

Yoder, P. Stanley, Robert Hornik, and Ben C. Chirwa, "Evaluating the Program Effects of a Radio Drama about AIDS in Zambia," *Studies in Family Planning*, Vol. 27, No. 4, 1996, pp. 188–203.

Acknowledgments

About This Report

The number of programs dedicated to countering violent extremism (CVE) has grown in recent years, but a fundamental gap remains in the understanding of the effectiveness of such programs. A 2017 RAND Corporation report documented that only a handful of such programs have been subject to rigorous evaluations of effect. Such evaluations are critical because they help ensure that programming funds are dedicated to the most-effective efforts. Evaluations also play a critical role in helping individual programs improve the quality of service provision.

This report presents the results of an evaluation designed to assess the impact of a CVE-themed radio program broadcast in northern Nigeria in 2018–2019. RAND researchers studied this program by recruiting more than 2,000 northern Nigerians via text message from a research panel administered by a mobile phone–based market research company. The participants were randomly assigned to listen to either the treatment program of interest or a nontreatment control program. This report details the research design and findings and offers recommendations for improving such evaluations in the future.

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

www.rand.org