MAX PLANCK INSTITUTE
FOR RESEARCH ON
COLLECTIVE GOODS

Felix Rusche

Discussion Paper
2025/5

BROADCASTING CHANGE: INDIA'S COMMUNITY RADIO POLICY AND WOMEN'S EMPOWERMENT

Broadcasting Change: India's Community Radio Policy and Women's Empowerment

Felix Rusche*

April 2025

Abstract

In poor countries, the interaction of early marriage, early motherhood, and low educational attainment disempowers women and limits their life opportunities. Even as countries grow richer, gender inequality is often sustained by social norms, thereby limiting welfare gains from women's empowerment. I investigate the use of media as a cheap and scalable policy to empower women. In 2006, India enacted a community radio policy that grants radio licenses to NGOs and educational institutions with the aim to foster local development. I collect original data on the content and coverage areas of all 250+ radio stations. I uncover women's empowerment as a key theme through topic modeling and ChatGPT-based analyses of radio show recordings. For identification, I exploit topography-driven variation in radio access and develop a novel econometric approach to deal with randomly displaced geolocated household data. The results show that women exposed to radio gain an additional 0.3 years of education and are 4.1pp (11%) more likely to obtain a secondary degree. In line with increased education, exposure reduces child marriages by 1.4pp (22%) and fertility of young women by around 10% while they are 11pp more likely to exhibit autonomy in household decisions. The findings demonstrate that community media can effectively address gender inequality.

Keywords: Mass Media, Policy, Women Empowerment, Spatial Econometrics, Education
JEL Classifications: O12, J13, J16, J18

Acknowledgments: I benefited from discussions and feedback from Camille Urvoy, David Yanagizawa-Drott, Rachel Glennerster, Yoko Okuyama, Ruben Enikolopov, Doug Gollin, Wladislaw Mill, Matthias Sutter, Randi Hjalmarsson , Philipp Ager, Natalie Bau, Yoshiyasu Rai, Michèle Tertilt, Suresh Naidu, David Atkin, Christina Gathmann, Karan Singhal, Michela Bia, Natalie Bau, Gianmarco León-Ciliotta, Elliot Motte, Christoph Rothe, Cristina Bellés-Obrero, Jörn-Steffen Pischke, Romain Wacziarg, Angelika Budjan, Franziska Scholl, Johannes Bauer, Patrick Sun, Johannes Gallé, Carsten Trenkler, Rana Cömertpay, Martin Fernandez Sanchez, Valentin Lang, Sebastian Vollmer, and Stefan Klonner. I am deeply grateful to Antonio Ciccone for his continued advice and support. The article also benefited from discussions with participants at ASSA/AEA Annual Conference 2024, RES Annual Conference 2024, Federal Communications Commission (FCC) Economic Seminar 2024, Europe Meeting of the Econometric Society 2024, Asia Meeting of the Econometric Society 2024, Nordic Conference in Development Economics 2023, German Development Economics Conference 2023, German Economic Association Annual Conference 2023, LISER, Universitat Pompeu Fabra, ETH Zürich, ifo Institute Munich, Max Planck Institute for Research on Collective Goods Bonn, University of Gothenburg, Bocconi University, South Asia Institute Heidelberg, Frankfurt-Mannheim Ph.D. Workshop, Barcelona School of Economics PhD Jamboree, CRC TR224 Annual Meeting and Young Researchers Workshop (Universities of Bonn & Mannheim), and the University of Mannheim. I thank Sven Wittig at Fraunhofer Institute for Telecommunications for his support in tuning the ITM algorithm. Finally, I thank Archana Kapoor, Astha Singh, Meenu Rawat, Alok Kumar, and others at SMART, Suboor Bakht at Heidelberg University's New Delhi bureau, Gaurishankar Kesarwani at MOIB, Frank Hoffmann, Stefan Klonner, and Johannes Gallé for greatly supporting my field research in India. **Funding:** Financial support by the German Research Foundation (DFG) through CRC TR 224 (project A04), Joachim Herz Stiftung, and the state of Baden-Württemberg through bwUniCluster/bwHPC is gratefully acknowledged.

* Felix Rusche, University of Mannheim & Max Planck Institute for Research on Collective Goods; felix.rusche@uni-mannheim.de

I Introduction

In poor countries, gender inequality is particularly stark as exemplified by the phenomenon of 'missing women' (Sen, 1992). Girls often obtain little education and tend to marry and have children at a young age. This severely constrains both the life opportunities of women and their contribution to economic development. On the other hand, the impact of economic development on women's empowerment is limited by persistent gender attitudes and norms (Duflo, 2012; Jayachandran, 2015). The media has been found to change attitudes and behavior when listeners can relate to stories and characters (DellaVigna and La Ferrara, 2015). However, evidence on translating the results on media and socioeconomic outcomes into a policy is largely confined to single-issue government campaigns or field experiments (e.g. Banerjee et al., 2019a; Khalifa, 2022).

In this paper, I investigate the use of media as a cheap and scalable policy instrument to empower women. In 2006, India enacted a new media policy to foster economic and social development. The policy enables educational institutions and NGOs to establish community radio stations to address local development issues via locally produced content. The radio stations are further barred from producing political news. Within this mandate, community radios make their own editorial choices.

I document that community radios established under the policy strongly focus on women's empowerment by analyzing radio show recordings using both topic models and GPT-based text analysis. To identify the effects on women's empowerment, I develop a novel econometric approach that allows me to combine randomly displaced geolocated household data with topography-driven variation in radio access. I find strong effects of exposure to radio on variables associated with women's empowerment. Women exposed to radio gain an additional 0.3 years of education and are 1.4pp (22%) less likely to marry while underage. Additionally, I find effects on women's fertility and autonomy.

My empirical analysis builds on original data on the location, coverage area, and content of all community radio stations established in India by 2020. The more than 250 radio stations cover an estimated 331 million people, representing 23% of India's population. Given radio stations' editorial freedom, I first investigate which development issues they focus on. For this, I scrape, transcribe, and translate over 5,000 audio recordings of radio shows and analyze their content using a topic model. I identify women's empowerment and education as key themes in radio programming. GPT-based content analyses reveal that the radio stations advocate for girls' education and family planning while opposing child marriage and domestic violence. This emphasis on women's empowerment and education does not solely reflect the radio operators' priorities. Rather, it is in line with a widely held view of development issues in India. According to India's 2005 World Values Survey, 35% of respondents view women's empowerment and 38% view education as among the country's most serious issues. Only poverty was cited more frequently.

These views hold across socioeconomic groups (Inglehart et al., 2014).

To causally identify the effects of radio on women's empowerment, I combine a well-established approach that exploits topographic variation between radio towers and listeners with a novel method to address the random displacement of geolocated household survey data. I use household survey data from the 2015-16 National Family and Health Survey (NFHS). The NFHS is part of the Demographic and Health Surveys, a repeated cross section conducted across most of the developing world. A major challenge when using this data is that, to protect respondents' privacy, coordinates of individuals' locations are randomly displaced (jittered) prior to being reported to researchers. Specifically, in rural areas, half of all observations are jittered by more than 2.5km. This is substantial relative to the variation in coverage driven by topographic features, with the average coverage area having a radius of around 15-25km. Consequently, relying solely on jittered locations leads to a mismeasurement of radio coverage when using the reported jittered location of households, which is the standard approach in existing research. Instead of relying on the reported locations, I develop a novel estimation approach that draws on the publicly known jittering algorithm. Based on the algorithm and high-resolution population data, I compute the probability density function of true survey locations given the reported location and combine it with the coverage areas of the community radio stations. I show that my approach can identify the effect of radio coverage on individual attitudes and behavior. As such, it can rule out multiple threats to identification inherent to using the standard approach. Most importantly, I can resolve the attenuation bias through a mismeasurement in the treatment variable. Following the same approach, key control variables, such as distance to the radio tower, are corrected, ruling out bias originating from mismeasurement in these.

I begin by examining whether individuals exogenously exposed to community radio are more likely to listen to radio broadcasts. I find that households exposed to community radio are 4.9ppt (25%) and 7ppt (42%) more likely to have heard a family planning message or message on HIV/Aids on the radio relative to the sample average. Hence, community radio stations increase exposure to messages they typically produce. I also find a positive, but somewhat weaker, effect on overall radio consumption. The effects are stronger for women, even though women generally listen to the radio less frequently than men (17% vs. 30%). Community radio stations thus reach an audience and, importantly, increase listeners' exposure to the type of messages they typically broadcast. I find no evidence for substitution away from other media, such as television, internet, or newspapers.

Given the strong focus of community radio on women's empowerment and education, I begin by examining the effects on educational outcomes. I focus my analyses on age groups potentially affected in their educational choices, that is, 5-30 year olds at the time of the survey. I find that exposure to community radio increases years of schooling, attendance rates, and the propensity to have obtained a degree. In my baseline specification, these effects are mainly driven by girls. The results indicate that living in an area exogenously exposed to radio makes young women around 3-4pp more likely to obtain a primary, secondary, and higher degree, respectively. Relative increases are strongest for higher (13%) followed by secondary (11%) and primary education (4%). Overall, total years of education of 5-30 year old girls and women, i.e. those who may have been affected by community radio, increase by around 0.3 years. Given the cumulative nature of the effect, effects are higher (0.5 years) when only focusing on 19 to 30 year olds, who are beyond secondary school age. Related survey responses on reasons for school dropout show that the findings are driven by an increase in the willingness to invest in girls' education, increased interest in school, and a decreased dropout rate due to early marriage. The effects are not driven by improved transport or school facilities for girls. The findings suggest changes in parents' aspirations for girls and possibly of girls' aspirations. They are not driven by supply side factors. For boys, I also find evidence of positive effects, although the effect sizes are smaller and the results are less robust. The results on education are consistent with a strong focus of community radio stations' on education and women's empowerment.

Community radio stations also impact the marriage market and fertility outcomes. Women exposed to radio are less likely to be married between the ages of 13 and 25. In relative terms, the effects on child marriage are particularly strong with a 1.4pp (22%) decrease for girls. Effects for men are lagged by around five years, likely due to the average five-year age gap between husbands and wives. For fertility, I find an 8-12% decrease in the number of children women have between the ages of 19 and 35. These findings may result from delayed child bearing due to later marriage or decreases in overall completed fertility. Given that most children are born when mothers are well below 35 years of age, a decrease in lifetime fertility is more likely. This reduction in fertility may also be in line with increased household bargaining power of women. Men generally have preferences for a higher number of children than their partners (Doepke et al., 2012) and women in my setting generally tend to get more children than they would find ideal.

Finally, I test whether community radio stations affect women's autonomy and variables on domestic violence. Young women, i.e., those most likely to have profited from additional education, are 11pp (21%) more likely to participate in household decisions or decisions about their own mobility. Tentative evidence suggests that men also adjust their attitudes toward the autonomy of women. Overall, they increase the share of decisions in which they believe women should participate by around 4.7pp (6.1%). I also find suggestive evidence of decreases in women's approval of domestic violence and the experience thereof. Male attitudes towards violence are unaffected.

Overall, my study provides evidence on community media as a powerful instrument to affect the role of women. The results suggest that community radios shift listeners' gender attitudes and behavior by producing messages on women's empowerment that are tailored to the community they serve. In addition to changes in revealed preferences as observed in education, marriage, and fertility, this is supported by evidence on attitudes and autonomy of women as well as men's beliefs regarding their wives' autonomy. In support of the findings on men's attitudes, I further find no evidence of a male backlash in terms of increased domestic violence. In fact, I find suggestive evidence of lower domestic violence in areas exposed to community radio. Changes in gender norms are also evident in shifting patterns in the reasons for school dropout. These suggest higher aspirations of parents regarding girls' education and higher aspirations of girls themselves. Overall, the findings can be interpreted as evidence in favor of a media-induced change in gender norms.

I rule out multiple potential threats to identification. A basic threat is heterogeneity in observables correlated with the treatment variation used for identification. Here, I examine the effect of exogenous topographic variation of radio coverage on socioeconomic and geographic variables that should not be affected by radio coverage (following Adena et al., 2020; Yanagizawa-Drott, 2014). The variables include scheduled caste/tribe shares, urbanity, population density, religion, and travel times to the nearest city, border, and radio tower. I find no effect of radio exposure on these variables. Hence, radio coverage does not appear to be related to a range of socio-economic or geographic observables. I also use two different approaches to examine heterogeneity in unobservables. First, I examine the effect of exogenous topographic variation of radio coverage on the education of age groups whose choices are unlikely to have been altered by radio. This includes cohorts that had already completed their education when the first radio stations were launched. I find no effects of radio exposure on the education of these age cohorts. This indicates that my results are not driven by time-invariant observables or differential trends. Second, I implement a placebo check that combines all my 2016 outcome variables with the coverage areas of community radios launched after 2016. Again, I find no effects on outcomes. This indicates that my results are unlikely to be driven by community radio being such that the coverage areas are related to the effects that I find. Finally, I vary the regression specifications in several different ways to ensure that the results hold up. Overall, the robustness and exogeneity checks suggest that neither heterogeneity in observables nor in unobservables drives the results, strengthening the results' causal interpretation.

I contribute to several literatures. First, I contribute evidence on the intended use of community media as a policy to the larger literature on media and socioeconomic outcomes (for reviews see La Ferrara, 2016; DellaVigna and La Ferrara, 2015; Enikolopov and Petrova, 2017). This literature can broadly be categorized into two main branches: first, a number of papers study the unintended effects of entertainment media using observational data (Kearney and Levine, 2015; Walsh, 2023). Most closely related, Chong and La Ferrara (2009), Jensen and Oster (2009), and La Ferrara (2016) document effects of the roll-out of entertainment media on women's empowerment as a result of exposure to different ways of life. A second strand of the literature tests the effectiveness of exposing individuals to specific movies or shows in field experiments (Arias, 2014; Bernard et al., 2014; Berg and Zia, 2017; Bjorvatn et al., 2020; Cassidy et al., 2022; Coville et al., 2019; Green et al., 2018; Green and Vasudevan, 2018; Kasteng et al., 2018; Murray et al., 2015; Ravallion et al., 2015; Riley, 2024). For example, Banerjee et al. (2019a,b) invite Nigerians to watch an MTV show featuring information on HIV and domestic violence. Despite this large body of research, there is very little evidence on the intended use of media as a policy instrument to affect socioeconomic outcomes aside from studies evaluating single-issue (government) campaigns (Glennerster et al., 2021; Khalifa, 2022).[1] A notable exception is Okuyama (2023)'s study on a radio program on women's status produced by the US occupying force in 1945-52 Japan. Based on district-level data, the author documents effects on political participation and fertility but none on education. My paper contributes by showing that grassroots media can be used as a policy instrument to affect the role of women at scale. The evidence I provide is the result of a policy passed by a developing country and through democratic processes. In addition, it is driven by hundreds of radios that operate within a common policy framework as opposed to a single station. Further, the radio stations do not produce content in distant urban areas, exposing listeners to different ways of life. Rather, they produce content within the community they serve. The evidence thus suggests a strong ability of grassroots media to adapt to local environments and the potential of such radio stations to affect outcomes. It also provides evidence on the effects of long-term exposure to mass media messaging as opposed to short-term exposure as is typical in field experiments. Not least, the data I use is at the individual-level as opposed to the district-level. This allows for a close examination of effects and mechanisms, including the evolution of effects over time and effects on radio consumption.

I also contribute evidence on the use of community media as a policy to the literature on campaigns and policies to empower women. As noted by two reviews on the relationship between economic growth and women's empowerment, policy intervention is required to achieve gender equality (Duflo, 2012; Jayachandran, 2015). Research on such interventions encompasses cash transfers (Baird et al., 2011), education subsidies (Duflo et al., 2015), adolescent training programs (Bandiera et al., 2020), inheritance reforms (Mookerjee, 2019), the elimination of school fees (Keats, 2018; Lucas and Mbiti, 2012), pension programs (Duflo, 2003) and others. I contribute by presenting evidence on the use of media as a policy instrument. As noted in a review on interventions to enhance women's agency by Chang et al. (2020), evidence on entertainment media sug-

[1] Qian (2024) further studies communist propaganda during China's cultural revolution and its effects on gender equality in education.

gests media interventions to be a promising path. However, the authors note the lack of evidence on large-scale interventions. I provide such evidence and show that community radio has strong effects on women's empowerment. Importantly, the policy effectively allows the government to draw on civil society's (i.e., NGOs' and educational institutions') knowledge and resources, rendering the policy cheap from the government's perspective.

I further relate to the literature on norms by providing evidence on media-induced changes in gender norms. As famously shown by Alesina et al. (2013), gender norms persist even as economic realities change. Moreover, gender norms might persist even if the majority of people updated their view, as people often underestimate the spread of such views (Bursztyn et al., 2020, 2023). In a literature review, Lowes (2022) notes that an important question is what changes culture and under what circumstances culture does change. Jayachandran (2015) cites the media as a potential pathway given evidence on the unintentional effects of entertainment media cited above. In my paper, I provide evidence on a media induced change in gender norms. In addition, I show that radios choose to strongly focus on women's empowerment in a setting with both pervasive gender inequality and a widely held belief of this being an issue (Inglehart et al., 2014).

Finally, I contribute methodologically by developing a novel econometric approach to estimate parameters in light of geographically displaced survey coordinates in Demographic and Health Surveys (DHS). With over 400 surveys across more than 90 countries, the DHS are likely the most widely used surveys in development economics and popular in other fields such as public health (Altay et al., 2022b).[2] Despite their widespread use, the standard approach in existing research is to simply use reported, that is, displaced, survey coordinates (Michler et al., 2022). As I document, this introduces substantial (attenuation) bias. Other research on this issue is scarce. To the best of my knowledge, the closest is Altay et al. (2022b), who suggest a way to compute prevalence maps of, e.g., educational attainment or disease, in light of the jittering.[3] More generally, my paper relates to the theoretical literature on identification under misclassification in treatment assignment (Hu, 2008; Mahajan, 2006; Schennach, 2016). Lewbel (2007) suggest an instrumental variable approach that instruments treatment assignment in the first stage. Unlike my study, the literature generally assumes no knowledge about the distribution of the misclassification error. I contribute by proposing and applying a novel method that can identify parameters when working with displaced survey coordinates in the DHS data.

[2] Using Google Scholar, I estimate that around 6.5% of studies published in the Journal of Development Economics and World Development as well as 117 articles in 'Top 5' journals use DHS data.

[3] Karra et al. (2020) further suggest a way to correct for the distance to health facilities when using displaced DHS locations. However, their paper ignores two important features of the DHS jittering algorithm and survey methodology, making the results inconsistent under standard conditions. As noted in their paper, it ignores reflecting administrative borders in the DHS and the differentiation between urban and rural areas. Therefore, it is only valid when it is used exclusively in locations that are not in the vicinity of a district border (>10km) and not close to an urban/rural area (>5km). This is very rarely the case in practice. In addition, they suggest that the approach they take to correcting for the bias may be computationally too expensive when adjusting for these additional factors.

In line with solving attenuation bias, the method increases point estimates by more than 50% on average. As such, the method proposed is relevant to studies using geocoded survey locations with a known displacement algorithm. Consequently, it lays the groundwork for future research, especially in situations where the displacement introduces excessive measurement error, rendering a given study design unviable.

II Context and Policy

II.1 Community Radio

Radio remains one of the most accessible media for people in developing countries. It is cheap, easily accessible including to illiterate populations, but is also easily translated into more modern media, e.g., through live streams or podcasts (UNESCO, 2013). The potential of radio to reach poor populations, led policy makers, activists, and international organizations to suggest the use of community radio for development (Fraser and Restrepo-Estrada, 2002; Raghunath, 2020). Community radio stations aim to offer marginalized communities a platform for addressing local concerns, promoting local customs and languages, and delivering information and education (Fraser and Restrepo-Estrada, 2002).

Although there is no comprehensive data on the global diffusion of community radio, many countries, especially across Africa and Latin America, have granted licenses to a large number of community radio stations. For example, 93% of the villages in northern Benin had access to at least one community radio stations in 2009 (Keefer and Khemani, 2016). Boas and Hidalgo (2011) count 2,328 stations in Brazil in 2008. In South Asia, where media is typically more strongly controlled by the state, community radio has only more recently started to gain pace (Raghunath, 2020).

India may be particularly suited to benefit from community radio. While adult literacy has increased, around a quarter of the adult population remains illiterate (World Bank, 2023). India is extremely diverse, both culturally and linguistically, with 122 languages and more local dialects (Census of India, 2002). Furthermore, a large part of the population lacks access to the media. In 2016, 15% of men and 25% of women reported not being regularly exposed to mass media, such as television, radio, cinema or newspapers (IIPS and ICF, 2017).

II.2 Community Radio in India: Policy

By the 1990s, India's state-run All India Radio (AIR) covered about 99% of the population. Frequently misused as a government mouthpiece (Kumar, 2003; Thomas, 2013), politicians were hesitant to give up control over airwaves until a 1996 supreme court ruling led to the first auctions of private FM licenses in 1999 (Kumar, 2003). These were focused on entertainment, cover around 45% of India's population, and are not allowed to broadcast news (KPMG, 2017) or even sexual education, rendering them "electronic discos for urban youth" (Fraser and Estrada, 2001, p. 28).

It took another decade of pressure from activists with the support of UNESCO for the government to pass legislation that allowed the establishment of community radio stations in 2006 (Pavarala and Malik, 2007). Compared to other countries, the regulation of community radio is quite restrictive with respect to the allocation of licenses and the content of radio programs. Starting with eligibility to apply for a license, three types of institutions can set up community radio stations: educational institutions, NGOs, and government-financed agricultural centers ('Krishi Vigyan Kendras') established to improve local agricultural practices (Varshney et al., 2022). Aside from these, neither individuals nor political organizations or commercial enterprises can receive a license. In addition, NGOs must be established for at least three years prior to submitting an application (Govt. of India, 2006).

To obtain a license, radios go through a rigorous licensing process. The process is conducted at the federal level meaning that local or state governments are generally not involved. There are two key bottlenecks in the application process, according to the director of the Ministry of Information and Broadcasting's community radio department, who I interviewed while conducting field research in India. These insights were confirmed by another interview with the head of a facilitator NGO that works with most community radio stations and applicants in India. First, many applicants fail to provide the necessary documents. Second, many applicants cannot convince the screening committee of their previous involvement with and connection to the community. The screening committee is led by the MOIB and, amongst others, comprises of community radio advocates, practitioners, UNICEF, and other stakeholders (Raghunath, 2020).

Once a station is set up, it is required to adhere to various content-related regulations. Importantly, the policy explicitly states that *"the emphasis should be on developmental, agricultural, health, educational, environmental, social welfare, community development and cultural programmes"* (Govt. of India, 2006, p. 5). At least half of this content must be produced locally and in a local language or dialect. The policy also prohibits radios from producing certain content. Importantly, it prohibits radio stations from broadcasting (political) news.[4] Further, it holds radio stations liable for spreading demeaning content about minorities and disadvantaged groups, such as women (Govt. of India, 2006).

To obtain funds, community radio stations can run 5 minutes of advertisements per

[4] Radios can, however, air newscasts produced by All India Radio (Myers, 2011) However, according to multiple expert interviews and interviews I conducted at five community radio stations, this only very rarely done in practice.

hour. In addition, they can apply for government funding for installation costs, participate in government communication schemes (CRFC, 2022) or seek funding from donors (Govt. of India, 2006).

III Data and Descriptive Statistics

III.1 Data Collection and Preparation

Community Radio Stations Data on community radio stations is collected from a variety of sources. First, a list of all 289 stations as of March 31, 2020 is obtained from the Ministry of Information and Broadcasting.Apart from the address and launch date, the list shows that 49% of all stations are run by NGOs, 45% by educational institutions, and 6% by Krishi Vigyan Kendra (KVK). Up to 2020, an average of 14 radios have been launched each year (also see Figure B.1 in Appendix).

I geolocated stations by rigorously searching for and identifying their precise location on the Web using information on their name, address, and license holder. Following my visit to the Ministry of Information and Broadcasting, the ministry also provided me with a list that includes approximate locations (1.2km precision), which I used to verify the collected information. In total, 276 of 289 stations were verified as operational of which 96% or 264 stations were precisely geocoded (see H in the appendix for further information on data collection and geocoding). Using the precise locations combined with information on radio tower height and transmitter power, radios' coverage areas are estimated using the Longley-Rice/Irregular Terrain Model.[5]

Merger with National Family and Health Survey The main data set for both controls and outcomes is the 2015-16 National Family and Health Survey (NFHS), India's arm of the DHS survey (IIPS and ICF, 2017). The data is representative both nationally and on the district level and includes information on 2.9 million individuals from 601 thousand households. Each of the 28k survey clusters includes around 21 households and is associated with unique coordinates.

I match NFHS cluster coordinates with estimated coverage areas of community radio stations. Given that many clusters are out of reach of any radio signal, I reduce the sample to observations with a realistic chance of being covered by a radio signal. In the paper's main specifications, this includes all observations at a distance of up to 50km from a radio tower. This includes 96% of the total coverage area.[6] Additionally, different

[5] Radio coverage areas are calculated using the ITM algorithm through `cloud.rf`'s API. Kasampalis et al. (2013) shows that the ITM model is highly precise, showing a correlation of 0.8 between estimated and actual coverage. Armand et al. (2020) validate this. In their setting, the correlation is even higher.

[6] Figure B.3 in Appendix.

thresholds are chosen as robustness checks.

I then create two separate and non-exclusive sets of observations: the main sample includes all observations within 50km of a radio that launched before 2016. This covers individuals whose outcomes may have been altered by the presence of a community radio station. The placebo sample, on the other hand, includes observations within 50km of a radio launched from 2016-20. Figure 1 shows the included and excluded data for the main sample.

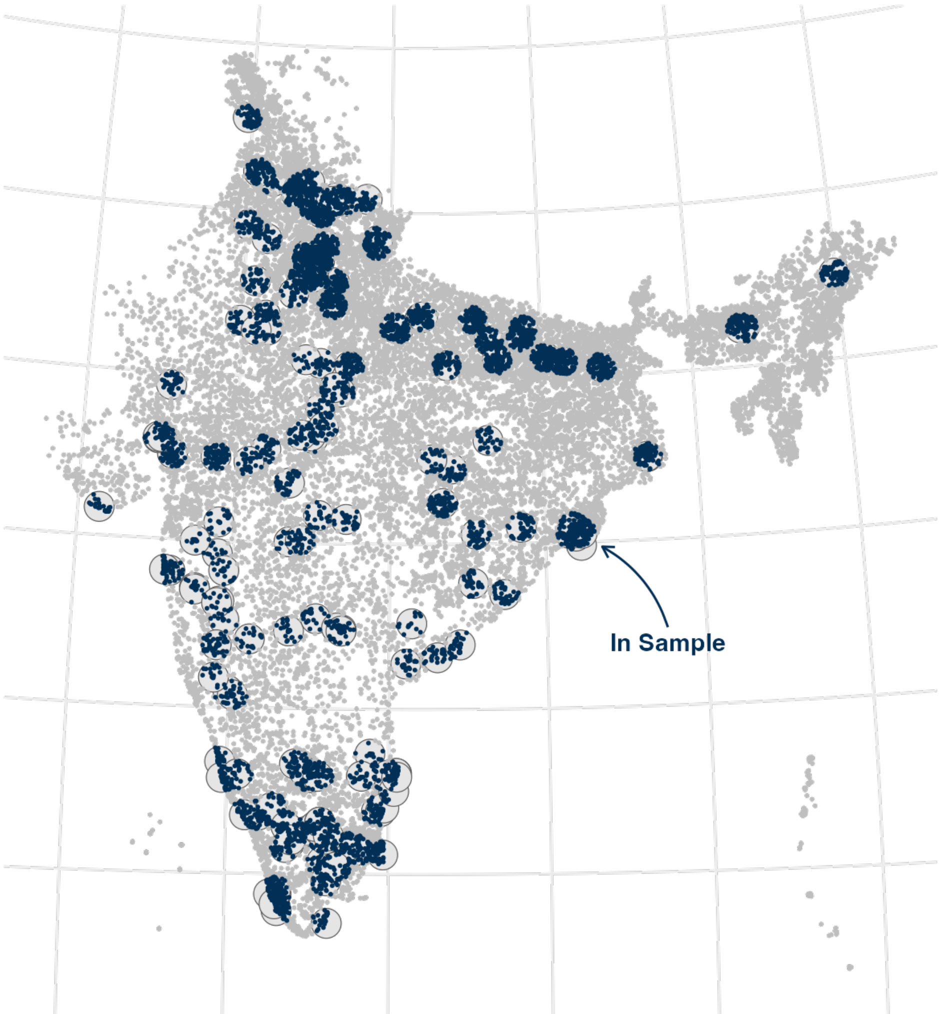


Figure 1: Visualization of NFHS data

Note: Observations within 50km of a given station are included, a total of 8,217 clusters. Each cluster includes around approximately 21 households.

Treatment and Outcomes Table 1 provides summary statistics of variables included in the regressions. In total, the data incorporates 821k observations from 8,217 clusters. All variables are reported at the individual level. Table B.1 (in the appendix) provides detailed descriptions of each variable and its source.

Starting with radio variables, the average probability of an individual being within the coverage area is 47%. Around 19% of individuals listen to radio at least less than once a week, and a similar number have heard a family planning message on radio in the past months. 16% report having received information on HIV/AIDS on radio. In total, only 9% of households possess a radio, suggesting that individuals often jointly listen to radio. For a developing country, these numbers are rather low, likely explained by the early and widespread introduction of television in India (Jensen and Oster, 2009).

The first set of outcomes refers to girls' and women's education, fertility, and marriage. Education is measured in three ways: first, through the number of years individuals spend in school. Second, by their highest earned degree, and third by whether a child is in school at the time of survey. Regarding fertility and marriage, women surveyed have an average of 1.7 living children while 72% have ever been married.

Autonomy describes a woman's ability to affect her life through own actions and decisions. It is an important mechanism through which women can alter their life prospects, including fertility and other outcomes (Jayachandran, 2017). I measure women's autonomy through their say in household decisions and with regard to their mobility. Importantly, outcomes on autonomy are only collected in about 35% of the survey clusters. Regarding mobility, women are asked whether they can visit different places alone, with someone else, or not at all. Three places are surveyed: the market, the health facility, and places outside the village. For decisions within the household, women are questioned about whether they make these decisions independently, together with their husband, or whether they are excluded from the decision-making process. Three decisions are surveyed: respondent health care, large household purchases, and visits to friends and family. As a measure of autonomy, I compute the share of places women can visit on their own and decisions they participate in. The variable therefore ranges from 0 to 1. The average suggests that respondents have autonomy with respect to 64% of decisions or mobility choices.

Similarly, the DHS also encompasses variables on men's standing towards their partners' or wives' autonomy. The questions posed to men differ in two aspects from those posed to female respondents. First, they ask about the respondent's views on who 'should have' rather than who 'factually has' a greater say with respect to different household decisions. Second, they only include variables regarding household decisions, i.e., none on mobility. I, again, code the variables such that a value of 1 means that a respondent believes that his wife should be involved in all household decisions.

Finally, I include outcomes related to domestic violence. Again, these outcomes are only collected in around 35% of survey clusters. Starting with attitudes toward violence, questions on whether women find it justified for husbands to beat their wives under specific circumstances are surveyed. These include arguing with husband, burning food, going out unannounced, neglecting children, and refusing sex. Following Jensen and

Oster (2009), I count the number of reasons for which a woman finds domestic violence justifiable. An alternative specification simply indicates whether the respondent finds domestic violence justifiable under any circumstance. Approximately 41% of women find domestic violence justifiable, with an average of 1.1 reasons mentioned. Notably, men report being less accepting of domestic violence. 30% agree with any reason for domestic violence. Finally, an even smaller sample of women is asked about their experiences with domestic violence. 33% of women experienced violence from their partner ever and 27% in the past 12 months.

Additional Variables A number of different groups of variables are included and are used as controls as described in Chapter V.1. The first set of variables pertains to demographics, including age, caste, religion, and sex. The second set of variables relates to variables affecting the propagation of radio signals. This includes the altitude and ruggedness surrounding survey clusters to altitude and ruggedness based on detailed elevation data provided by Jarvis et al. (2008).[7] Propagation controls further include the (expected) distance to the nearest radio tower (also see Chapter IV). In addition, I compute the travel time from each observation to the nearest radio tower using Google's Direction API.[8] Finally, additional geographic controls cover the urbanization, population density, travel times to the nearest city, and distances to water bodies and national borders.[9]

III.2 Descriptives: Content

Depending on the audience and aim of the institution running a particular radio station, community radio stations focus on a host of different issues. The role of women has been a leading cause of activists fighting both for the policy and of operational community radio stations. Pavarala and Malik (2007) summarize that *"gender is a significant dimension in community radio initiatives that are seeking to deploy communication technologies for social change in general and empowerment of women in particular"* (p.210). Overall, women are not only addressed as an audience, but also strongly involved in the management structure and content production of many community radio stations (Pavarala and Malik, 2007; Nirmala, 2015).

The first source I use to explore radio content are 'Community Radio Compendia'. These booklets have regularly been published as part of 'CRS Sammelan', a facilitator event for community radio stations. They provide a one-page fact sheet on each partici-

7 Specifically, I compute the average altitude and ruggedness within the 5km surrounding the reported location, hence, following the DHS' practices for computing geographic controls.

8 The data is visualized in Figure B.6 (in Appendix).

9 Travel times to the nearest city are based on Weiss et al. (2018). They define a high-density urban area "[...] as a contiguous area with 1,500 or more inhabitants per square kilometer or a majority of built-up land cover coincident with a population centre of at least 50,000 inhabitants" (p.333).

pating station, including a short description of the radio's main focus area and content. For radios that did not participate, the information is enriched with information from radio stations' websites (if available). In total, I collect content information on 248 radios.[10] After identifying the main topics, I go through all the texts, manually marking words related to different topics (see Section I in the appendix for more information on the procedure and underlying data). Overall, 129 or 54% of radios explicitly mention words related to 'women empowerment' in their self-description, making it one of the most common themes. Education is mentioned by 64% of radio stations. Other key topics are health & hygiene, culture, and agriculture and fishing.

The widespread coverage of topics related to women empowerment are confirmed by a survey of 160 radios conducted by SMART, an NGO working with community radios in India. It shows that 90% of surveyed stations broadcast programs related to gender and *"the majority of community radios are broadcasting programs on child marriage, sexual harassment, gender-based violence, and women and health education"* (p. 4). The survey also shows that more than half of all staff members are women, who particularly work in content production and as radio jockeys (SMART, 2023).

To gain a clearer understanding of the topics discussed by radio stations, I crawl all >14k radio shows uploaded to `edaa.in`, a platform where community radios can upload and exchange content.[11] The shows were uploaded from 2011 to 2019 with the median show having been uploaded in early 2014. Using Google's Speech-to-Text API and Google Translate, I transcribe and translate 5,869 shows from 95 stations that uploaded content to the website.[12] After cleaning the transcripts, Latent Dirichlet allocation (LDA) is applied to identify topics (Blei et al., 2003). LDA is arguably the most widely used method for determining latent topics in a selection of documents. Intuitively, it treats each transcript as a mix of latent topics, where topics are probability distributions over terms. Each document is assumed to have been created by drawing from the distributions of these topics. Based on these assumptions, the terms, and the chosen number of topics, LDA estimates the topic distribution for each document and the term distribution for each topic (Hansen et al., 2018, provide a detailed description of LDA, including its underlying econometrics).[13] The resulting topics are hand-labeled based on each topic's 15 most predictive terms (see Table J.4 in the appendix).

To get an idea of the content, I first collapse the topics into 8 categories. The graph

[10] Of these, information on 211 radios stems from radio compendia. Thereof, 180 descriptions are from the 2019 version.

[11] The website appears not to be used by listeners, as exemplified by the fact that Google Trends does not rate the website on the vast majority of days in the period of interest.

[12] Given that some radios uploaded a host of content, I randomly choose up to 578 shows from the radios that uploaded more than that.

[13] The transcription and translation of audio files naturally reduced the resulting transcripts' quality. Although the process retains words used, it often does not retain sentence structures. For this reason, I decided against using topic models that take into account context and sentences.

on the left of Figure 2 visualizes the average radio's share of development-related content across topics.[14] As visible, radios cover a lot of ground, ranging from agriculture to education and women-specific content. In addition, the topics of women and education make up around half of the total content. The right-hand side of the figure further zooms into women-specific topics. These include subtopics on women's health, education, maternity, and marriage.

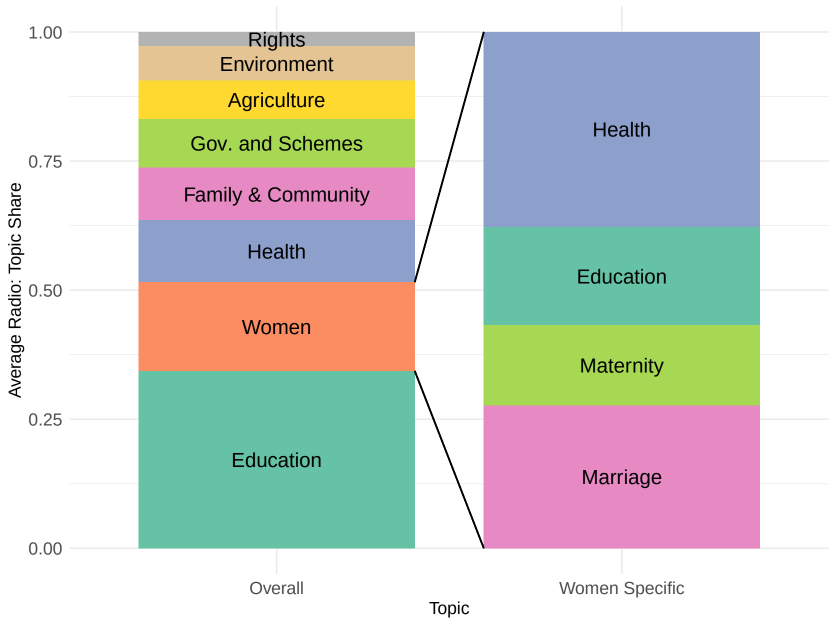


Figure 2: Radios' share of content across topics based on LDA

Note: The above figure visualizes the distribution of topics of the average radio station. For this, translated transcripts of radio shows are assigned topic shares using an LDA model. Next, the average transcript is computed by station. Finally, the average radio's content is computed. I exclude entertainment and undefined other topics from the visualization in order to provide an idea of development-related messages.

To empirically test what stance radios take on topics related to women's empowerment, I employ a novel approach to prepare and analyze radio shows through a multistage evaluation using Generative Pre-Trained Transformers (GPT). I start by preparing the transcripts for analyses using GPT. The translation and transcription process strongly affects the grammatical structure and interpretability of transcripts. To prepare these for content analyses, I first send all transcripts to OpenAI's ChatGPT-3.5 requesting a restoration of the original transcript without adding any additional information or making assumptions. After preparing the transcripts, I classify whether these discuss the topics of child marriage, girls' education, family planning, or violence against women. Specifically, I ask GPT-4 to return a vector with four binary variables indicating whether the respective topic is discussed. Similarly to using multiple research assistants, the request is sent twice and, in case the two answers are in disagreement, a third request is sent, applying a majority rule. I identify potential additional articles on the topics using

[14] I removed content from the 'entertainment' topic as the focus of the analysis is on development related content.

simple keywords, such as 'child marriage' or 'contraception'. In a final step, I then send all identified transcripts to ChatGPT-4o. I first ask whether the article covers the respective topic and, if so, ChatGPT is asked to state whether the articles are in favor, neutral or against the respective issue (e.g. child marriage or girls' education). Chapter J.2 in the appendix provides a detailed explanation of the approach, including specific prompts.

The results show that 96% of the points of view taken on the above issues can be described as in line with women's empowerment in the sense that they argue in favor of girls' education and family planning as well as against child marriage and domestic violence. In total, 387 or 6.6% of shows are identified to explicitly discuss the issues listed above. These take 423 viewpoints in favor of women's empowerment, as defined above.[15] Only two opposing points of view are identified, with another 18 taking no or a neutral position. Overall, this suggests that the content produced by radio stations can be described as in line with what would usually be considered women's empowerment.

In general, each of the three sources concerning radio content comes with its own set of strengths and weaknesses. For example, radio stations may selectively upload shows or report selective topics in radio compendia. They may also shift their focus over time. However, taken together, the sources consistently demonstrate that women empowerment and education are vital elements of community radio stations' content. Additionally, they demonstrate that the content aligns with traditional definitions of women's empowerment found in research, such as advocating for girls' education, delayed marriage, reduced fertility, and opposing domestic violence.

[15] This number is slightly higher than the number of shows given that some shows discuss multiple issues.

Variable	# Survey Clusters	# Individuals	Mean	SD	Median	Min	Max
Radio Variables							
Exposure	8,211	821,243	0.23	0.27	0.10	0	0.99
Coverage Probability	8,211	821,243	0.47	0.44	0.33	0	1
Coverage Probability: Closest Radio	8,211	821,243	0.44	0.44	0.26	0	1
Radio Owner	8,207	195,584	0.09	0.29	0	0	1
Radio Consumer	8,208	234,550	0.19	0.39	0	0	1
Radio Familyplanning	8,208	234,550	0.20	0.40	0	0	1
Radio HIV/AIDS	2,844	56,782	0.16	0.37	0	0	1
Outcomes							
Years of Edu.	8,210	819,532	5.90	5.25	5	0	20
Completed Primary	8,210	819,532	0.50	0.50	0	0	1
Completed Secondary	8,210	819,532	0.18	0.38	0	0	1
Higher than Secondary	8,210	819,532	0.11	0.31	0	0	1
Ever Married	8,210	620,620	0.72	0.45	1	0	1
Num. Children	8,207	202,106	1.70	1.61	2	0	15
Autonomy of Women (Female Respondent)	2,842	24,983	0.64	0.33	0.67	0	1
Autonomy of Women (Male Respondent)	2,843	31,231	0.82	0.29	1	0	1
Attitude (Count)	2,842	34,188	1.10	1.61	0	0	5
Attitude (Any)	2,842	34,188	0.41	0.49	0	0	1
Attitude (Count) - Male Respondent	2,842	31,628	0.70	1.28	0	0	5
Attitude (Any) - Male Respondent	2,842	31,628	0.30	0.46	0	0	1
Experienced Violence by Partner (Ever)	2,839	18,825	0.33	0.47	0	0	1
Experienced Violence by Partner (Past 12m)	2,839	18,825	0.27	0.44	0	0	1
Controls: Demography							
Age	8,210	821,138	29.33	20.11	26	0	95
Female	8,210	821,242	0.49	0.50	0	0	1
Caste ST	8,192	799,207	0.22	0.42	0	0	1
Caste: SC	8,192	799,207	0.07	0.25	0	0	1
Caste: OBC	8,192	799,207	0.46	0.50	0	0	1
Caste: Other	8,192	799,207	0.25	0.43	0	0	1
Religion: Hindu	8,210	821,242	0.81	0.39	1	0	1
Religion: Muslim	8,210	821,242	0.14	0.34	0	0	1
Religion: Other	8,210	821,242	0.06	0.23	0	0	1
Controls: Propagation							
Travel Time to Radio Tower (min)	8,194	819,525	57.28	33.89	55.18	0.75	329.83
Distance to Radio Tower (km)	8,211	821,243	26.00	14.71	26.86	0.91	49.99
Distance to 2nd closest Tower (km)	8,211	821,243	67.99	58.09	53.23	1.35	433.91
Mean Altitude	8,211	821,243	274.85	300.71	209.30	-0.06	2,471.05
Mean Ruggedness	8,211	821,243	10.95	17.66	5.70	2.24	156.25
Geographic							
Urban	8,210	821,242	0.39	0.49	0	0	1
Pop. Density (2015)	8,211	821,243	2,509.62	5,915.53	857.46	23.24	63,807.06
Travel Time to Nearest City (min)	8,211	821,243	14.58	17.38	11.24	0	275.48
Proximity: Water (m)	8,211	821,243	177,990.70	118,651.60	174,432.00	1.96	511,661.20
Proximity: National Borders (m)	8,211	821,243	180,751.90	130,304.30	159,207.20	10.38	583,496.30

Table 1: Summary statistics: DHS

Note: The table above provides summary statistics of the paper's main sample, that is, survey clusters at a distance of up to 50km around a radio tower launched prior to 2016. The table shows both the number of survey clusters (as shown in Figure 1 and the number of individuals. Appendix Table B.1 provides further details on the respective variables, including their sources.

IV Defining the Treatment Variable and Correcting for Spatial Jittering

Accurate information on treatment is essential for identification. In this paper, treatment is primarily defined by residing inside a radio's coverage area. For individuals' locations and demographic outcomes, the best available data source is the Demographic and Health Survey, namely India's 2015-16 NFHS. As part of the survey, enumerators gather precise coordinates of each enumeration area. Such areas are small geographic units that cover around 20 households each. However, to ensure the privacy of the respondents, the precise coordinates of enumeration areas are jittered by up to 2km in urban and 5km in rural areas.[16] Through this, a substantial measurement error is introduced into the variable describing treatment assignment. Despite this, previous research has largely ignored the jittering and simply relied upon jittered coordinates.

I propose a way forward by developing a method to compute the expected probability of treatment conditional on the observed jittered location. To be more precise, I draw upon public knowledge of the jittering algorithm and, in an application of Bayes' law for random variables, compute the probability density function (PDF) of original locations conditional on observing a jittered location. In a second step, I combine the PDF with the coverage area to compute the probability mass on the coverage area. Using the same approach, I further compute the expected distance between radio towers and individuals. As I will argue in Section V.2, the proposed method produces consistent estimates of the treatment effect in the presence of jittering.

IV.1 The Jittering

After collecting coordinates of enumeration areas, the DHS applies the following jittering algorithm which follows the "random direction, random distance" method (for a detailed description see Burgert et al., 2013):[17]

1. randomly choose an angle between 0 and 360 degrees with uniform distribution

2. randomly choose a distance according to the type of cluster (urban: 0-2km / rural: 0-5km) with uniform distribution across the distance

3. combine both draws to obtain a new coordinate

As a result, the PDF of the jittering algorithm resembles a 'circus tent'. The algorithm further has one important exception: if the jittered location drawn above lies outside the

[16] Further, in rural areas, 1% of clusters are displaced by up to 10km. Given that, in expectation, only 0.5% of the clusters are jittered by more than 5km, this part of the jittering process is ignored here, as it has virtually no impact on estimation while substantially increasing computational costs when estimating coverage.

[17] For a formalization of the displacement see Altay et al. (2022a).

administrative unit, a new location is drawn until the draw results in a location within the given administrative unit. In India, administrative units refer to districts.
The PDF of drawing jittered location x conditional on survey location x^* can be characterized as:

$$f(x|x^*) = \frac{I(A(x^*) = A(x)) \times I(d(x^*, x) \leq \bar{d})}{d(x^*, x)} / \underbrace{\int_z \frac{I(A(x^*) = A(z)) \times I(d(x^*, z) \leq \bar{d})}{d(x^*, z)} dz}_{w(x^*)} \tag{1}$$

$$= \begin{cases} \frac{1}{d(x^*,x)} \times \frac{1}{w(x^*)}, & \text{if } A(x^*) = A(x) \text{ and } d(x^*, x) \leq \bar{d} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$A(.)$ describes the administrative unit of a given location, d(.,.) describes the distance between two locations, and $\bar{d}$ describes the maximum jittering distance, i.e., 2km in urban and 5km in rural areas. As shown, for valid locations x, the PDF depends on two components: $d(x^*, x)$, and $w(x^*)$. Importantly, $w(x^*)$ can be understood as the share of the full 'circus tent's' distribution lying within the administrative boundaries of $A(x^*)$. This means that for valid x, $f(x|x^*)$ increases for locations in the vicinity of a border. To see this, consider two locations x_1^* and x_2^*. x_1^* is far from the border and its jittering PDF follows a circus tent. x_2^* is just next to a straight border. Here, the circus tent is cut in half. As a result, the probability weight on any viable location x doubles as $w(x_1^*) \approx 2 \times w(x_2^*)$.

Information on the jittering algorithm can then be used to obtain information on the PDF of original locations x^* conditional on observing a jittered location x. Specifically, Bayes' theorem for random variables, states:

$$f(x^*|x) = \frac{f(x|x^*)f(x^*)}{\int_z f(x|z)f(z)dz} \tag{3}$$

From Equation 1, $f(x|x^*)$ is well defined for any x^*. $f(x^*)$ describes the distribution of original survey locations across space or, more specifically, the 2/5km radius around x.

There are two viable approaches which, at least in my setting, yield very similar results. The first option is to estimate $f(x^*)$ using empirical data. Specifically, the DHS generally differentiates between urban and rural areas. Within either, the survey locations are chosen with probability proportional to their population. Specifically, within the same adminstrative unit (district) and within urban/rural areas of that unit, the PDF follows the population distribution (DHS, n.d.). I draw on two data sets to compute just that, namely the 100m resolution population data from WorldPop (2020) and Balk et al. (2019, 2020)'s urban/rural definitions of India's 2011 population census at 1km resolution which the 2015-16 NFHS follows (IIPS and ICF, 2017).

The second option is less data hungry, but requires the assumption that original locations are uniformly distributed across space (within viable administrative units). For observations at a distance of more than $2 \times \bar{d}$ from the border the PDF has the nice feature of simplifying to: $f(x|x^*) = f(x^*|x)$, i.e. suggesting that the distribution of original survey locations follows the jittering distribution. It does, however, forego gains from placing more weight on areas with a high population density, such as on villages surrounded by agricultural land. While the first approach is empirically preferable, I discuss the second approach for illustrative purposes and for applications in settings without detailed data on urban/rural definitions.

In general, the above formula can then be combined with the treatment area to estimate the probability mass of original locations on the treatment area conditional on observing location x:

$$\Pi(T = 1|x) = \frac{\int_z T(z) \times f(x|z) \times f(z)dz}{\int_z f(x|z) \times f(z)dz} \tag{4}$$

where $T(z)\epsilon\{0, 1\}$ is the treatment status of location z.

Computationally, the above can be be implemented as follows:

For each jittered and observed location x:

1. create an equidistant grid of points z within distance $\bar{d}$ and administrative unit $A(x)$
2. only when estimating $f(x^*)$: reduce z to those in rural/urban areas, depending on x being reported as rural or urban
3. only when estimating $f(x^*)$: compute the population living at z as a share of the population living across all valid z: $p(z) = population(z)/\sum_z population(z)$
4. for each z: generate a second equidistant grid with points v at distance $\bar{d}$ of z and in adminstrative unit $A(x)$ and compute $\sum_v \frac{1}{d(z,v)}$
5. again, only for viable z and v, estimate:

$$\Pi(\widehat{T = 1}|x) = \frac{\sum_z [T(z) \times \overbrace{\frac{1}{d(z,x)} / \sum_v \frac{1}{d(z,v)}}^{\widehat{f(x|z)}} \times \overbrace{p(z)}^{\widehat{f(z)}}]}{\sum_z [\frac{1}{d(z,x)} / \sum_v \frac{1}{d(z,v)} \times p(z)]} \tag{5}$$

(leave out $p(z)$ when assuming a uniform distribution over space)

The result $\Pi(\widehat{T = 1}|x)$ is what I, for simplicity, term 'coverage probability': the probability mass on original locations located within the treatment area, conditional on observing a jittered location at coordinate x.

Figure 3 visualizes Equation 5. Part (a) shows the different elements of the equation. The figure in the upper left shows the location of an observed location x as well as the district border and the coverage area. The top right displays the population distribution, that is $\widehat{f(z)}$. It also further shows that no weight is put on locations outside the district or on urban areas given that the reported location is rural. The third figure displays the fact that for locations close to the district border, the probability of drawing any valid location within the district is higher. Note that given that boundaries between urban and rural areas within the same district are not reflective, the pattern is not observed at such boundaries. Finally, the bottom right figure of Part (a) shows the simple distance weight, showing that locations closer to x have a higher likelihood. Part (b) then combines the resulting estimates and shows the probability density function of original locations conditional on observed location x. Finally, Part (c) visualizes the likelihood mass on the coverage area. While the reported location is inside the coverage area, only 66% of its probability mass does.[18]

IV.2 Jittering and Expected (Squared) Distances

The jittering does not solely affect the measurement of treatment status. It also affects the calculation of distances between observed locations x and other points of interest, such as the radio tower.

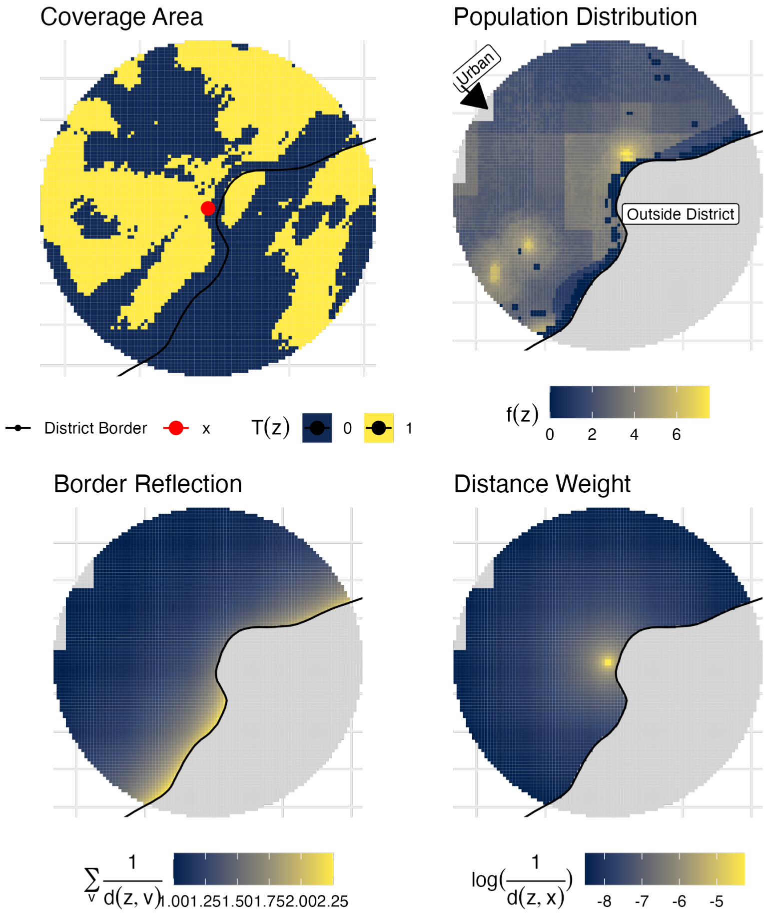
The distance between an observed location x and a radio tower t generally does not equal the expected distance between x^* and t conditional on x:

$$\mathbb{E}(d(x^*, t)|x) = \int_z f(z|x)d(z, t)dz \tag{6}$$

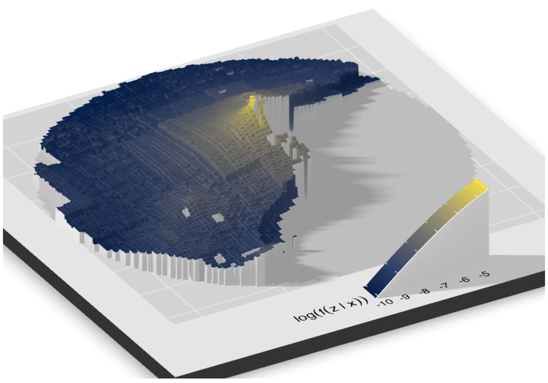
The computation can be performed using the box provided above by computing the distance between z and t in Step 4 and replacing $T(z)$ by $d(z, t)$ in Step 5.

It is also possible to obtain a closed-form solution for the above equation. This applies to locations not affected by a district border and when assuming $f(x^*)$ to be uniform. To see this, first note that given the uniform distribution assumption $f(x^*|x) = f(x|x^*)$. In other words, the PDF of original locations exactly follows the jittering algorithm. Further note that jittering is uniform in direction and distance. Thus, if one were to split the 5km circle around the reported location into two 'donuts', one going from 0 to $0.5\bar{d}$ and the other going from $0.5\bar{d}$ to $\bar{d}$, each donut contains the same probability mass of original locations. Similarly, when drawing a large number of donuts with the same width, each contains the same probability mass given uniform jittering across the distance. Now consider drawing one of these donuts at distance r from x and a potential

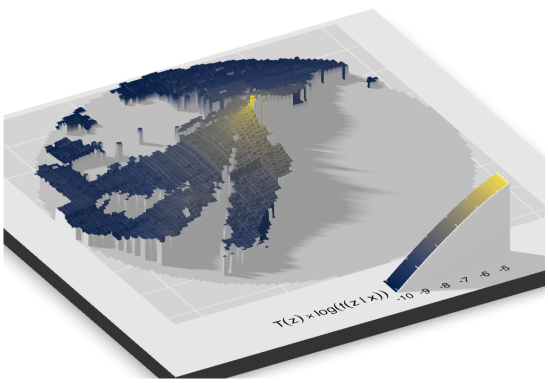
[18] Figure A.1 visualizes the object assuming $f(x^*)$ to be uniform and x not being close to an administrative border.



(a) Visualization of elements of Equation 5



(b) PDF of original survey locations $log(f(z|x))$



(c) PDF on coverage area $T(z) \times log(f(z|x))$

Figure 3: Visualization of computation of coverage probability

Note: Part (a) shows the ingredients of Equation 5. Part (b) shows the resulting likelihood mass on original locations. I rescale the mass using the log for illustrative purposes. Finally, Part (c) shows the likelihood mass on the coverage area. The coverage probability of this example location is 66% and the reported location x is covered. The figures show the 100x100m grid reaching 5km around the reported location x. The figures are created using ggplot2 and rayshader (Morgan-Wall, 2024; Wickham, 2011).

original location z that lies in this donut as shown in Figure 4. To calculate the distance $d(z,t)$ between z and a radio tower t, one can draw on the Law of Cosines, which states $d(z,t) = \sqrt{r^2 + d(x,t)^2 - 2rd(x,t)cos(\phi)}$. Integrating the distance formula for a uniformly distributed variable $\phi \varepsilon [0, 2\pi)$ then provides the expected distance between the point t and any z on the circle. Intuitively, this moves z in infinitesimal steps once around the circle. At each step, $d(z,t)$ is calculated and, overall, its expectation. Finally, integrating over all donuts between 0 and 2/5km yields the expected distance between x^* and t:

$$\mathbb{E}(d(x^*,t)|x) = \frac{1}{\bar{d}} \int_0^{\bar{d}} \frac{1}{2\pi} \int_0^{2\pi} \sqrt{r^2 + d(x,t)^2 - 2rd(x,t)\cos(\phi)} d\phi dr \tag{7}$$

$$= \frac{1}{\bar{d}} \int_0^{\bar{d}} \frac{2}{\pi} (r + d(x,t))^2 EI(\frac{2\sqrt{rd(x,t)}}{r + d(x,t)})\, dr \tag{8}$$

Equation 8 follows by re-writing the equation as a function of the elliptic integral of the second kind $EI(.)$, which allows for efficient calculation of the expected distance as a function of r and $d(x,t)$ (see Appendix A for a detailed derivation).

Comparing the expected difference between x^* and t to the reported one between x and t yields several insights: First, as shown in Figure A.2 (in the appendix), by the above formula, no location is expected to be at a distance below 2.5km from the radio tower (1km in urban areas). This is true, even if x is exactly equal to t. To see this, note that even if this were the case, the original location lies anywhere between 0 and $\bar{d}$ from the observed location with uniform probability across the distance. Therefore, x^* is expected to be at a distance of $\frac{1}{2}\bar{d}$ from the tower. The figure also shows that this insight broadly remains after taking into account district borders and population weighting. Second, the absolute and relative difference between $d(x,t)$ and $\mathbb{E}(d(x^*,t)|x)$ decreases in $d(x,t)$. However, $d(x,t)$ is always smaller than $\mathbb{E}(d(x^*,t)|x)$. Third, given that the difference increases in $\bar{d}$, differences for urban areas are smaller.

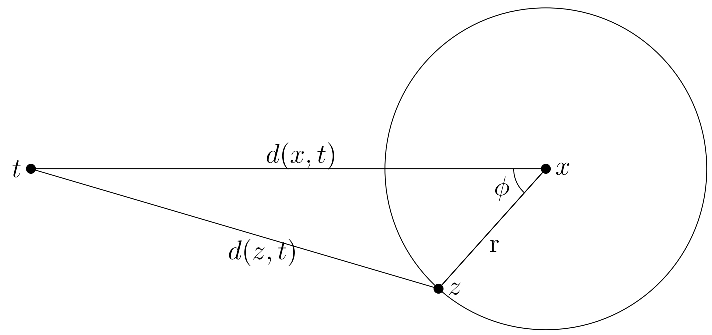


Figure 4: Visualization of setup and Law of Cosines

Finally, the expected squared distance can be derived following the same logic as above, namely, by additionally computing the squared distance between z and t in Step 4 and replacing $T(z)$ by $d(z,t)^2$ in Step 5. Making the above assumptions, again allows for the derivation of a closed form solution:

$$\mathbb{E}(d(x^*,t)^2|x) = \int_0^{\bar{d}} \frac{1}{2\pi} \int_0^{2\pi} r^2 + d(x,t)^2 - 2rd(x,t)cos(\phi)\, d\phi dr \tag{9}$$

$$= \begin{cases} d(x,t)^2 + 8.\bar{3} \text{ in urban clusters} \\ d(x,t)^2 + 1.\bar{3} \text{ in rural clusters} \end{cases} \tag{10}$$

The above equation shows that under the simplifying assumptions, the expected squared distance only varies between urban and rural clusters, i.e., by how far locations are jittered. Thus, other than for the expected distance, the difference between the expected squared distance and $d(x,t)$ is a constant number.

Overall, the results regarding expected (squared) distances suggest that studies controlling for distances between DHS observations and any geographic object or border should correct for these, especially when working on rather small geographic areas or when distances are vital controls, such as distances to treatment areas or locations.

V Empirical Strategy

V.1 Identifying Effects of Community Radio Stations

To identify the causal effect of community radios, variation in coverage due to local topographical features is exploited (Olken, 2009).[19] This is done in several steps: First, using the irregular terrain model (Hufford, 2002) and with information on the power, location, and height of the radio transmitter as well as the topography of India, the coverage area of each community radio station is estimated.[20] Given that the location of the transmitter may be correlated with other unobservable characteristics, e.g., if radios tend to be built in more or less developed areas, controls for the distance to the transmitter are included (Armand et al., 2020; Yanagizawa-Drott, 2014).[21] The remaining variation in radio coverage is driven by differences in the line of sight between the transmitter and the observation. This is affected by both the topography between the observation and the transmitter, as well as the topography of the observation's immediate surroundings.

[19] The strategy or variations thereof are used in a number of papers, e.g. Adena et al. (2020, 2015); Armand et al. (2020); Bursztyn and Cantoni (2016); DellaVigna et al. (2014); Enikolopov et al. (2011), and Yanagizawa-Drott (2014)

[20] The height is officially restricted to be between 15 and 30m. However, multiple expert interviews at NGOs and the ministry and visits to four radio stations confirmed that radios maximize their coverage by building a 30m tower.

[21] an alternative is to control for the theoretical radio signal received by the observation in free space (e.g. see Durante et al. (2019) and Olken (2009)). It does not fit the context of this paper, given that it is unclear how to define the coverage probability in such a setup.

The latter may directly affect outcomes, for example, because places up in the mountains may be less likely to receive the signal and be more conservative. To control for this, topography controls are added. These include second-order polynomials of the altitude and ruggedness of observations' surroundings. Finally, I control for the time it takes to travel to the radio tower. This additional control directly captures both the distance to the closest radio tower and the geographic surroundings of specific locations. The topography between the radio tower and the receiver, i.e. household, drives the remaining variation in coverage.

To account for level differences between different parts of India, radio fixed effects are added, where each observation obtains the fixed effect of the closest radio station (that was launched before data collection). The resulting estimator exploits the variation in received radio signals within such areas.

Given that radios launch at different points in time (see Figure B.1), the potential effects of radio do not solely depend on their presence at the time of data collection but also on how long they have been on the air. Thus, even if an individual lives right next door to a community radio station, the radio is not expected to have any effect if it is launched a day before data collection. Following the logic of Armand et al. (2020), treatment is thus defined as follows:

$$Exposure_{c(i)} = \sum_{r=1}^{R} AddedCoverageProbability_{c(i),r(i)} \times f(Timeshare_{r(i)}) \tag{11}$$

where $f(Timeshare_{r(i)})$ is a function of the share of time between 2005 and 2015 that radio $r(i)$ was on air where $Timeshare_{r(i)}$ ranges from 0 to 1. $AddedCoverageProbability_{c(i),r(i)}$ describes the increase in probability to be covered by a radio signal that radio $r(i)$ brings (ranging from 0 to 1) in addition to previously launched radios. In general, around 80% of observations are only covered by the station closest to their location. Further, 93% of the coverage probability stems from the closest station.

It is generally unclear what functional form the effects take over time. The true functional form may also differ from outcome to outcome. For example, outcomes dependent on factual information may immediately change people's behavior, while messaging that questions social norms may have weaker effects at first that grow stronger over time. To take an agnostic approach, I assume linear effects over time in my baseline specification. I further explore alternative functional forms in Chapter C in the appendix. The results suggest that the effect on radio consumption may be nicely resembled by a quadratic effect over time. I hence also provide estimates for quadratic effects over time, i.e., setting $f(Timeshare_{r(i)}) = Timeshare^2_{r(i)}$.

Moving the above into a regression framework yields the following specification:

$$y_i = \beta Exposure_{c(i)} + DistanceControls_{c(i)}\delta + GeoControls_{c(i)}\omega + X_i\lambda + \gamma_{r(i)} + \epsilon_{i,c(i),r(i)} \quad (12)$$

here, y_i is the outcome of interest for individual i. $DistanceControls_{c(i)}$ includes second-order polynomials of the expected distance to the closest, second, and third closest radio towers for i's cluster $c(i)$.[22] Further, I control for the travel time between the cluster and the closest radio tower.[23] $GeoControls_{c(i)}$ includes topography controls, i.e. second-order polynomials of ruggedness and altitude. To more precisely control for local development, X_i includes several variables related to clusters' location and demographics. The following variables are included: log. population density, log. travel time to the nearest urban area as defined by Weiss et al. (2018), proximity to national borders and water bodies, whether the cluster is defined as urban by the NFHS (follows the 2011 Indian population census' definition of urban/rural), age dummies, caste (ST/SC/OBC/Other), religion (Hindu/Muslim/Other), gender, and an interaction between urbanity and gender to account for general differences in women empowerment between urban and rural India (e.g. Asher et al., 2021; Biswas and Banu, 2023). Finally, $\gamma_{r(i)}$ are fixed effects for radio r closest to individual i. This controls for level differences across treatment areas. Finally, $Exposure_{c(i)}$ is the variable of interest, which describes the exposure of cluster $c(i)$ to community radio. A value equal to 1 suggests that the location has been covered with full probability over the entire time period, that is from 2005 to 2015.

Identification relies on exogenous variation in exposure to community radio stations driven by topographical features between the radio tower and the observation. Although the treatment variable includes the share of time a radio was present in a given region, it is important to note that $\gamma_{r(i)}$ effectively controls for any specific characteristics of the coverage area. This includes the fact that certain areas receive a community radio stations at an earlier point in time. Thus, identification is based on topographical features. Specifically, the identification assumption is that the remaining variation of exposure is driven by topographical features between the transmitter and the receiver and uncorrelated with all other determinants of women's empowerment.

All regressions are estimated using OLS. In line with Armand et al. (2020) and Yanagizawa-Drott (2014), I account for spatial autocorrelation using Conley (1999, 2010) Standard Errors with a 100km spherical kernel. In addition, the main results are estimated using heterosketasticity robust standard errors clustered at the subdistrict level (see Section G in the appendix). This follows Durante et al. (2019), DellaVigna et al. (2014), Adena et al. (2020), and Olken (2009).[24]

[22] Values for the second and third closest towers are capped at 50km.
[23] See Figure B.6 for a visualization
[24] Key packages used: Regressions: *fixest* (Bergé, 2018); Spatial operations: *sf* (Pebesma, 2018); Table Export: *modelsummary* (Arel-Bundock et al., 2023) (all in R)

V.2 Uncovering Regression Parameters in Presence of Jittering

In this subsection, I show that the estimating equation can recover the true parameters. I also discuss the bias arising without the proposed correction.

First, for simplicity, I define a stylized version of the above equation. I am generally interested in estimating the following equation:

$$y = \beta_0 + \beta_1 I(x^*) \times T + \beta_2 d(x^*) + Z\gamma + \varepsilon \tag{13}$$

$$with \quad \mathbb{E}[\varepsilon | I(x^*) \times T, d(x^*), Z] = 0 \tag{14}$$

where $I(x^*) \times T$ is the exposure variable, $d(x^*)$ is the distance to the radio tower, and Z includes additional controls unaffected by the jittering. Similarly, the time since the radio station's launch T is unaffected by the jittering. Importantly, x^* is the true survey location that I do *not* observe.

Simply replacing x^* with the observed location x, as would be the standard approach in current research, introduces measurement error in the treatment variable. If no other variables were affected, this would result in simple attenuation bias, hence downward biasing treatment effects. Given that, as shown above, estimates of the distance are affected as well, the issue potentially gets more severe as it is unclear which direction the bias takes.

In my study, I propose to instead estimate the above equation using the conditional expectation of y given jittered location x and controls Z:

$$\mathbb{E}[y|x, Z] = \beta_0 + \beta_1 \underbrace{\mathbb{E}[I(x^*)|x, Z]}_{\text{Coverage Probability}} \times T + \beta_2 \underbrace{\mathbb{E}[d(x^*)|x, Z]}_{\text{Expected Distance}} + Z\gamma + \mathbb{E}[\varepsilon|x, Z] \tag{15}$$

Importantly, I explicitly compute the conditional expectations of both $I(x^*)$ and $d(x^*)$ as described in Section IV.

The consistency of the above estimator thus hinges on whether the conditional expectation of ε equals zero. This follows when slightly adjusting the OLS assumption made in Equation 14. Specifically, adjusting the assumption to $\mathbb{E}[\varepsilon|x^*, Z]$ ensures that both equation 14 holds (by Adam's rule) and $\mathbb{E}[\varepsilon|x, Z] = 0$ holds. The latter follows from:

$$\begin{aligned}
\mathbb{E}[\varepsilon|x, Z] &= \mathbb{E}[\mathbb{E}[\varepsilon|x, Z, x^*]|x, Z] && ||x = x^* + \mu \qquad (16)\\
&= \mathbb{E}[\mathbb{E}[\varepsilon|x^* + \mu, Z, x^*]|x, Z] && ||x^* \perp \mu \qquad (17)\\
&= \mathbb{E}[\mathbb{E}[\varepsilon|\mu, Z, x^*]|x, Z] && ||\varepsilon \perp \mu \mid x^*, Z\\
&= \mathbb{E}[\underbrace{\mathbb{E}[\varepsilon|Z, x^*]}_{=0}|x, Z] = 0
\end{aligned}$$

Here, the first step follows from the fact that x is a function of x^* and the random jittering μ. The second step follows from the jittering being random noise and, hence, independent of x^*. This means that given x^*, I can uniquely determine μ. Similarly, the third step, again, follows from μ being random noise and, hence, being conditionally independent of ε.

The adjusted identification assumption states that the expected value of ε is zero for every combination of true location x^* with observables Z. This is slightly more abstract than the standard assumption in Equation 14 which states that the expected value of ε is zero for every combination of treatment status, distance, and observables, the former of which are a function of x^*. It is, however, in line with previous studies on radio. Specifically, these are typically conducted at the village, district or cell level, measuring average exposure to radio of a geographic unit. The studies therefore make the assumption of no relationship between the unobserved distribution of population characteristics within a given geographic unit and exposure to radio. In line with that, previous studies have run regressions to test whether the variation they use for identification is correlated with observables that should not be affected by the said variation (e.g. Adena et al., 2020; Yanagizawa-Drott, 2014).

I follow the same logic and test whether the variation I am using for identification is correlated with location characteristics that are unlikely to be affected by the radio but likely predictive of outcomes. That is, I test for heterogeneity in observables of the location being correlated to the treatment variation used in my paper. For this, I regress various characteristics on the treatment variable. The regressions slightly differ from Equation 12 by excluding variables in X_i, which partially serve as outcomes here. The regressions reported in Table D.13 are insignificant. This holds across different specifications, i.e., when assuming either linear or quadratic effects of radio over time. Overall, this strengthens the causal interpretation of the variation used in this paper. In addition, it is important to note that all the outcomes used in the exogeneity regressions are included as controls in the regressions below.

Panel A: Linear Effects Over Time

	Caste SC/ST	Muslim	Urban	Log. Pop. Density	Log. Travel Time City (min)	Proximity Borders (m)	Travel Time Radio (min)
exposure	−0.014	0.026	0.054	0.311	0.033	817.310	3.585
	(0.022)	(0.019)	(0.079)	(0.249)	(0.191)	(2873.507)	(2.688)
Num.Obs.	167 086	171 878	171 878	171 878	171 878	171 878	171 878
R2 Adj.	0.075	0.100	0.363	0.793	0.586	0.985	0.825
Distance Controls	✓	✓	✓	✓	✓	✓	✓
Geography Controls	✓	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Caste SC/ST	Muslim	Urban	Log. Pop. Density	Log. Travel Time City (min)	Proximity Borders (m)	Travel Time Radio (min)
exposure2	−0.037	0.005	0.022	0.306	0.059	421.987	2.903
	(0.025)	(0.022)	(0.088)	(0.266)	(0.225)	(2864.688)	(2.783)
Num.Obs.	167 086	171 878	171 878	171 878	171 878	171 878	171 878
R2 Adj.	0.075	0.100	0.363	0.792	0.586	0.985	0.825
Distance Controls	✓	✓	✓	✓	✓	✓	✓
Geography Controls	✓	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓	✓

Table 2: Correlation of treatment variation with observables

Note: The table show regressions of different covariates that are unlikely to be affected by radio on radio exposure. Regressions control for geographic and distance controls as well as radio fixed effects only. Regressions on travel time to the nearest radio station additionally exclude this variable from the set of distance controls. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

VI Results

In this section, I present the results for the effects of community radio on various outcomes. I first show that community radio increases radio consumption and, importantly, consumption of content typically produced by community radio. I then explore effects on variables associated with women's empowerment, including education, marriage, and fertility. Following this, I test effects on attitudes and autonomy of women. Finally, I discuss various placebo and robustness checks and evaluate the jittering correction.

VI.1 Radio Consumption and Content Reception

I start by investigating whether exposure affects radio consumption, including the consumption of development-related content. Table 3 reports the results. Starting with Column (2), being fully exposed to the radio from 2005 to 2016 is estimated to increase radio consumption by 3.0pp. in the linear and 4.8pp. in the quadratic model, though the results for the linear model do not rise to conventional levels of significance. Overall, the reported coefficients correspond to an increase of 16 to 26% percent compared to the baseline, that is, the mean of the dependent variable.[25]

[25] Table D.2 (in the appendix) shows that around half of the effect is driven by additional people rarely listening to radio, i.e. 'less than once a week'. The other half is driven by additional daily or weekly

More importantly, Columns (3) and (4) provide evidence on development-related content. These variables get closest to measuring exposure to content typically produced by community radio stations, as suggested by the content analyzes above. Specifically, the survey includes questions about having heard messages related to family planning or HIV/AIDS on the radio in the past months. The results show strong increases across these variables, ranging from 4.9 to 7.8pp. depending on variable and model. This suggests strong increases by 25 to 48% compared to baseline when fully exposed over the entire period of time. Separately estimating coefficients for men and women further reveals stronger effects for women, especially regarding family planning messages (Table D.1 in the appendix).

Finally, Column (1) shows that the observed effects are not driven by increased radio ownership. This may be unsurprising given that (in cash-constrained settings) it is rather unlikely for individuals to purchase a radio due to the arrival of a single additional radio station. As the difference between the ownership rate and consumption indicates, people also listen to radio jointly. This is likely to be the case for community radio stations that attempt to bring communities or specific groups, such as women, together.

Finally, Table D.12 (in the appendix) tests the effects of community radio stations on other media. On the one hand, this resembles a flawed robustness check, as one would not expect exposure to have a strong positive effect on other media. It is flawed in the sense that one may expect negative coefficients if people substitute other media for listening to radio. The results provide no such evidence. Exposure is not related to watching television, reading newspapers, using the Internet, or mobile phones. This is reassuring of the exogeneity check in Table D.13 and suggests that people do not stop consuming other media to listen to radio.

Overall, the results show that community radio increases radio consumption and strongly increases individuals' propensity to have listened to development-related messages via radio. These effects are not driven by substitution away from other media.

VI.2 Education, Marriage, and Fertility

Moving to key variables related to women's empowerment, the effects on three interrelated variables are investigated: education, marriage, and fertility. Aside from labor market effects, education has been linked to reduced fertility (Heath et al., 2024). Lower fertility, on the other hand, reduces the health risks of women due to birth and increases incentives to invest in women's human capital (Jayachandran and Lleras-Muney, 2009; Basu, 2002). It also frees up women's time, which can be spent on breaking out of traditional roles, e.g. by acquiring more education or participating in the labor market (Miller, 2010; Goldin, 2006). Similarly, child or early marriage generally limits women's potential to accumulate

listers.

Panel A: Linear Effects Over Time

	Radio Owner	Radio Consumer	Radio Familyplanning	Radio HIV/AIDS
exposure	−0.012	0.030	0.049**	0.070***
	(0.017)	(0.019)	(0.024)	(0.026)
Num.Obs.	190 090	228 215	228 215	55 484
R2 Adj.	0.065	0.073	0.098	0.092
Mean Y	0.095	0.184	0.197	0.164
SD Y	0.293	0.388	0.397	0.37
Controls	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Radio Owner	Radio Consumer	Radio Familyplanning	Radio HIV/AIDS
exposure2	−0.006	0.048**	0.060**	0.078***
	(0.016)	(0.020)	(0.026)	(0.029)
Num.Obs.	190 090	228 215	228 215	55 484
R2 Adj.	0.065	0.073	0.098	0.092
Mean Y	0.095	0.184	0.197	0.164
SD Y	0.293	0.388	0.397	0.37
Controls	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓

Table 3: Exposure and radio consumption

Note: The table shows the regression of radio consumption related variables on exposure. Regressions include all controls mentioned in Chapter V.1. The dependent variables are defined as follows: radio owner: household owns a radio; radio consumer: dummy indicating whether individual listens to radio at least less than once a week; radio family planning: dummy for whether individual heard a family planning message in last few months. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

skills and human capital (Giacobino et al., 2024). Early married girls often drop out of school before or once they get married (Jayachandran, 2015) and aspirations to marry daughters at an early age reduces parents' investments in girls (Maertens, 2013).

VI.2.1 Education

Starting with education, three variables are available: first, I estimate effects on years of education obtained across school types. Second, I estimate effects on the degree obtained. Finally, I test the effects on school attendance at the time the survey was conducted and changes in reasons cited for girls dropping out of school.

I start by evaluating the effects on years of education obtained. For this, I first define age groups that correspond to the age at which individuals are typically in lower primary (5-10), upper primary (11-14), lower secondary (15-16), upper secondary (16-18), and higher education (19-30) (Anderson and Lightfoot, 2019). Given that the underlying data constitute a cross section, effects regarding years of education are potentially additive between school types, as educational choices may have been altered at earlier stages of their school life. Furthermore, since the first radios launched around 10 years before the data was collected, no effects on education of individuals above the age of 30 are expected, who are likely to have completed their educational choices when the first radios launched.

Table 4 provides estimates on the education of boys and girls in the respective age cohorts. Strong effects on girls' and positive though lower effects on boys' education are shown. The latter are insignificant in regressions with linear effects over time and significant when allowing for quadratic effects. The impact on education increases between age groups and most strongly so when moving to upper primary, lower secondary, and higher education. Increasing coefficients in general suggests that effects are present in schools of all types.

Panel A: Linear Effects Over Time

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-19)	Higher Education (19-30)	All (5-30)
Female x exposure	0.062	0.204**	0.382**	0.299*	0.513**	0.310***
	(0.064)	(0.094)	(0.162)	(0.177)	(0.214)	(0.116)
Male x exposure	0.039	0.180**	0.114	0.117	0.212	0.174
	(0.059)	(0.079)	(0.119)	(0.180)	(0.195)	(0.116)
Num.Obs.	91 320	62 569	31 699	45 384	174 331	392 228
R2 Adj.	0.637	0.345	0.195	0.186	0.233	0.534
Mean Y	1.68	5.941	8.346	9.66	9.458	6.996
Mean Y: Female	1.706	5.999	8.341	9.557	8.902	6.821
Mean Y: Male	1.656	5.887	8.35	9.755	10.033	7.164
SD Y	1.635	2.116	2.584	3.406	4.912	4.855
Controls	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-19)	Higher Education (19-30)	All (5-30)
Female x exposure2	0.064	0.247**	0.483***	0.537**	0.800***	0.477***
	(0.066)	(0.112)	(0.164)	(0.221)	(0.237)	(0.137)
Male x exposure2	0.053	0.250***	0.117	0.405**	0.573***	0.377***
	(0.056)	(0.080)	(0.110)	(0.198)	(0.196)	(0.123)
Num.Obs.	91 320	62 569	31 699	45 384	174 331	392 228
R2 Adj.	0.637	0.345	0.195	0.186	0.233	0.534
Mean Y	1.68	5.941	8.346	9.66	9.458	6.996
Mean Y: Female	1.706	5.999	8.341	9.557	8.902	6.821
Mean Y: Male	1.656	5.887	8.35	9.755	10.033	7.164
SD Y	1.635	2.116	2.584	3.406	4.912	4.855
Controls	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓

Table 4: Effects of community radio on years of education by age group

Note: The tables show separate regressions of years of education by age cohort on exposure to radio. Panel A assumes linear effects over time and Panel B quadratic effects. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Next, I investigate whether exposure to radio affects the degree completed. Other than India's school system, the NFHS only differentiates between three types of degrees: primary, secondary, and 'higher'. I restrict the sample to those who have had the opportunity to obtain the respective degree and whose choices may have been altered by the community radio station. Based on NFHS data, this includes people between the ages of 15 and 30 for primary and 18 to 30 for secondary or higher education.

The results shown in Table 5 suggest that exposure to radio over the period of interest increases the probability for girls of obtaining a degree by 3-4pp across school types. Using a squared effect over time suggests similar results and effects for boys as well. In general, the findings confirm the additive interpretation of the results in Table 4.

Moving to school attendance, Table D.11 (in the appendix) tests effects on the propensity of a child to be in school at the time of the survey. This information is only collected for individuals between the ages of 5 to 18 and therefore does not cover higher education. The results suggest an increase in attendance in lower secondary and, in particular, higher secondary education.

To answer why girls obtain more education, I draw on information on the reasons for which 5-18 year olds drop out of school (Tables D.3 and D.4 in the appendix). Specifically, I create indicator variables for different reasons for dropout. I set the variables to zero for those still in school or those who have dropped out for another reason. The results show that falls in dropout rates are primarily driven by three factors. First, fewer students report a loss of interest in school as a reason for dropout. This particularly applies to girls and boys in lower secondary school. Costs are also substantially less likely to be cited as a reason for dropout for girls in higher secondary school. Finally, fewer girls in upper secondary school drop out due to marriage. Interestingly, reasons that primarily pertain to girls, such as safety, the lack of female teachers, lack of a school for girls, or household and care work, do not drive lower dropout rates. Similarly, work as a reason for dropout is unaffected.

Overall, results on education suggest strong effects on girls' and lower, frequently insignificant effects on boys' education. Although the results vary slightly by outcome and specification, the picture that emerges is consistent with additive effects across school types. This means that the propensity for kids to obtain additional education increases at all levels of education. The results on years of education and school attendance further suggest that the effects are strongest for secondary and higher education and weakest for lower primary education. This is consistent with the fact that while a large share of students finish primary school, secondary and higher education are the key barriers at which girls especially tend to drop out (Anderson and Lightfoot, 2019). The reasons for school dropout further suggest that effects are driven by increased interest in school, higher willingness to pay, and a decrease in the propensity of girls to enter an early marriage. This is indicative of an overall higher value placed on girls' education, changes in parents' aspirations for their daughters, and potentially girls' own aspirations.

VI.2.2 Marriage

Moving to the marriage market, Figure 5 provides evidence on the effects of radio by age group and sex. Across all regressions, the dependent variable describes whether an individual has ever been married. As before, coefficients arise from separate regressions by age group. The results show that the propensity for a woman to marry decreases up to her mid-20s, including decreases in early marriage between the ages of 13 to 18. I include 18 into 'early marriage' given that the age describes the age at which the survey took

Panel A: Linear Effects Over Time

	Primary	Secondary	Higher
Female x exposure	0.032**	0.041**	0.032**
	(0.015)	(0.020)	(0.014)
Male x exposure	0.013	0.021	0.020
	(0.013)	(0.020)	(0.015)
Num.Obs.	238 339	191 824	191 824
R2 Adj.	0.161	0.160	0.134
Mean Y	0.807	0.41	0.255
Mean Y: Female	0.766	0.383	0.241
Mean Y: Male	0.848	0.437	0.269
SD Y	0.395	0.492	0.436
Controls	✓	✓	✓
Radio FE	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Primary	Secondary	Higher
Female x exposure2	0.051***	0.067***	0.047***
	(0.016)	(0.023)	(0.014)
Male x exposure2	0.031**	0.063***	0.056***
	(0.013)	(0.021)	(0.016)
Num.Obs.	238 339	191 824	191 824
R2 Adj.	0.161	0.160	0.134
Mean Y	0.807	0.41	0.255
Mean Y: Female	0.766	0.383	0.241
Mean Y: Male	0.848	0.437	0.269
SD Y	0.395	0.492	0.436
Controls	✓	✓	✓
Radio FE	✓	✓	✓

Table 5: Effects of community radio on level of education achieved

Note: The dependent variable indicates whether an individual has obtained this degree, including individuals that obtained a higher degree. The results are presented for individuals aged 15-30 for primary and 18-30 for secondary and higher education at time of data collection (2015-16). These age groups are chosen as their choices may have been affected by community radio stations and given that they have been able to finish the respective degree. Panel A assumes linear effects over time and Panel B quadratic effects. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

place rather than the age of marriage. The result confirms the above results on marriage being less frequently cited as a reason for school dropout. Although early marriage results are low in absolute terms, they are high in relative terms. The point estimate of the linear model suggests a 22% decrease in the average of the dependent variable for girls/women when exposed over the entire time period. At 8.5%, the relative effect for women between the ages of 19-24 is lower but remains high. As would be expected, the propensity to have been married changes for men as well. In all, the point estimates are similar in magnitude but lagged by around 5 years. Men's marriage rates decrease most strongly between the ages of 25 and 29. This is consistent with an average age gap between husbands and wives of approximately 5 years in my data. By the age of 30 to 34, coefficients return to zero. At this age, most individuals in the sample are married, with little difference to

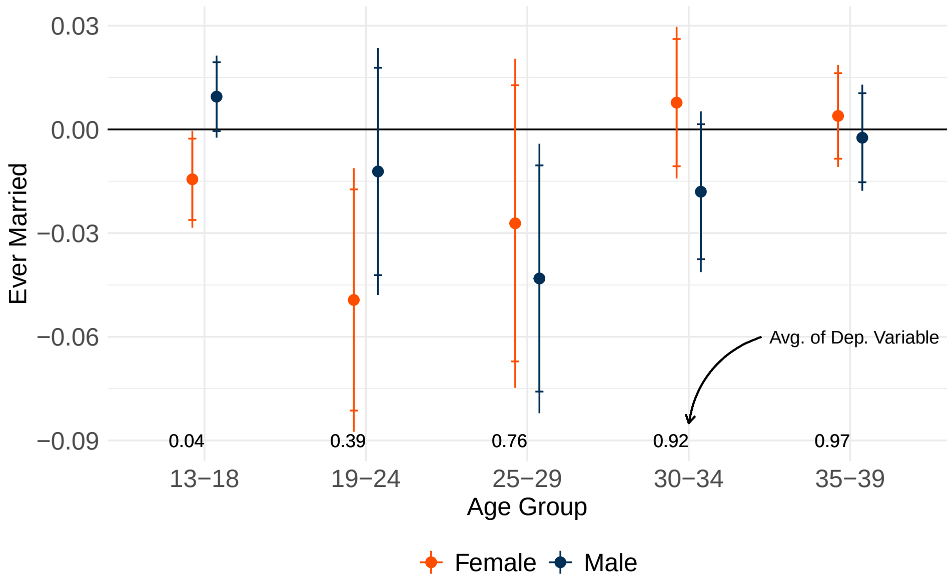


Figure 5: Effects of community radio on marriage

Note: The figure shows coefficients with 90 and 95% Confidence Intervals of regressions of a dummy for being married interacted with gender on radio exposure. Regressions are run separately by age group. These include all controls mentioned in Chapter V.1. Full regression results are shown in Table D.5. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010).

the overall marriage rate beyond this age (92% for 30-34 and 97% for 35-39 year olds). Table D.5 (in the appendix) provides the full regression results and confirms the above results using the quadratic model.

Overall, these findings suggest that exogenous exposure to community radio results in substantial delays in marriage, including early marriage of girls.

VI.2.3 Fertility

Table 6 presents the results with respect to the fertility of women. More specifically, it shows the number of children of women, both in general and by age group. The findings indicate that exposure to radio over the entire period of interest reduces the number of children by 0.1. Effects are particularly strong for individuals between the ages of 19 and 35. In absolute terms, effects are strongest for women aged 31-35, while there are no effects for older cohorts. The strong effect might be explained by older individuals having had more time to both have and not have children. Decreased fertility might be driven by both delayed child bearing due to later marriage or decreases in total lifetime fertility. Given that most children are born when mothers are well below 35 years of age, a decrease in lifetime fertility appears a more likely explanation.

Overall, the results with respect to fertility, marriage, and education suggest strong effects of radio exposure on women's status. In particular, educational choices can be interpreted as changes in attitudes toward girls' education while education is - in itself - an important mechanism to increase women's agency (Basu, 2002). Delayed marriage and reduced fertility provide further evidence of a change in the role of women.

Panel A: Linear Effects Over Time

	# Children (15-18)	# Children (19-25)	# Children (26-30)	# Children (31-35)	# Children (36-40)	# Children (41-49)
exposure	−0.003	−0.077**	−0.145**	−0.204**	−0.022	−0.031
	(0.003)	(0.033)	(0.069)	(0.094)	(0.076)	(0.102)
Num.Obs.	20 742	56 819	32 494	26 462	24 898	35 057
R2 Adj.	0.011	0.306	0.198	0.232	0.255	0.282
Mean Y	0.006	0.625	1.883	2.429	2.735	2.992
SD Y	0.084	0.899	1.199	1.295	1.429	1.633
Controls	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	# Children (15-18)	# Children (19-25)	# Children (26-30)	# Children (31-35)	# Children (36-40)	# Children (41-49)
exposure2	−0.002	−0.099***	−0.187***	−0.307***	−0.147	−0.151
	(0.003)	(0.038)	(0.066)	(0.082)	(0.110)	(0.094)
Num.Obs.	20 742	56 819	32 494	26 462	24 898	35 057
R2 Adj.	0.011	0.306	0.199	0.232	0.255	0.282
Mean Y	0.006	0.625	1.883	2.429	2.735	2.992
SD Y	0.084	0.899	1.199	1.295	1.429	1.633
Controls	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓

Table 6: Fertility: number of children

Note: The tables show separate regressions the number of children a woman has by age cohort on exposure to radio. Panel A assumes linear effects over time and Panel B quadratic effects. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

VI.3 Autonomy and Attitudes

While the above findings suggest improvements in women's autonomy and status, this section extends the analysis to attitudes toward domestic violence and women's autonomy. Survey responses on these are only collected in around a third of survey clusters, meaning that the treatment variation available for identification is substantially reduced (from around 8k to 3k clusters).

Panels (A) and (B) of Table 7 present results on autonomy, where the dependent variable is the share of decisions a woman participates in and the places she is allowed to visit on her own. Here, a value of 1 signifies that the respondent participates in all household decisions and can visit any surveyed location on her own. The results show overall positive effects driven by young women between the ages of 15 and 25. These are driven by increased autonomy in both decisions and women's mobility (Tables D.6 and D.7 in the appendix).

Panel (C) and (D) show a shift in men's views on their wife's involvement in household decisions as well. Other than for women, the coefficients are positive across all age groups up to the age of 45. However, only two coefficients are significant at the 5 or 10% level, hence results should be taken with a grain of salt.

Table D.8 (in the appendix) provides further evidence on attitudes. Specifically, it shows regressions on whether women find it justifiable for their husbands to beat their wives. The results are congruent with those on reported autonomy in the sense that coefficients suggest decreases in approval of domestic violence, especially among younger cohorts. However, coefficients do not rise to conventional levels of significance and should, hence, be interpreted with caution. The results for men suggest no change in attitudes (Table D.9 in the appendix).

Finally, Table D.10 tests the effects of exposure on experience of any sexual, physical, or emotional violence from a female respondent's partner. In line with the results on women's attitudes toward domestic violence, point estimates suggest a reduction, and more strongly so for younger cohorts. This is driven by decreases in the experience of physical rather than sexual or emotional violence. However, most coefficients are insignificant, rendering the results rather suggestive.

Overall, I document increases in young women's autonomy and men's attitudes toward women's autonomy. The results further suggest improvements in women's attitudes toward and experiences of domestic violence. However, given the small sample size, results on domestic violence are rarely significant and should be interpreted with caution. The results do, however, suggest no 'male backlash' against improvements in female empowerment. The absence of a 'male backlash' might be attributed to the nature of community radio, which disseminates information and perspectives from within the community itself. Hearing peers on the radio may make it less likely for backlash to occur. The fact that men's views become more favorable toward women's autonomy underlines the idea that their views are also altered by community radio. The potential of peer effects being activated by community radio may therefore have advantages compared to social change originating outside the community (e.g. Guarnieri and Rainer, 2021).

VI.4 Robustness and Placebo

In this section, I discuss robustness and placebo checks. A first potential threat to identification is heterogeneity in observables related to the treatment variation I am using for identification. Table D.13 suggests that this does not drive the results.

A second potential threat is heterogeneity in unobservables. I test for this in two ways. First, I repeat regressions on school degrees and years of education for individuals that have likely finished their educational choices by the time the first radios arrived. Specifically, I repeat regressions for individuals above the age of 30. These were aged 20 and above when the first radios launched. Tables F.1 and F.2 (in the appendix) show no effects of exposure to radio on these cohorts' educational outcomes. As a second test, I repeat all main regressions on a placebo sample. The placebo sample includes observations in the vicinity of a radio station that launches after data collection, that is,

Panel A: Women - Linear Effects Over Time

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-49)
exposure	0.036*	0.108***	0.013	0.006	0.060
	(0.020)	(0.034)	(0.022)	(0.029)	(0.060)
Num.Obs.	24 400	5481	9571	7210	2138
R2 Adj.	0.147	0.137	0.097	0.091	0.105
Mean Y	0.635	0.505	0.639	0.704	0.717
SD Y	0.329	0.336	0.322	0.309	0.308
Controls	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓

Panel B: Women - Quadratic Effects Over Time

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-49)
exposure2	0.019	0.106**	−0.003	−0.022	0.058
	(0.026)	(0.047)	(0.028)	(0.028)	(0.068)
Num.Obs.	24 400	5481	9571	7210	2138
R2 Adj.	0.147	0.136	0.097	0.091	0.105
Mean Y	0.635	0.505	0.639	0.704	0.717
SD Y	0.329	0.336	0.322	0.309	0.308
Controls	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓

Panel C: Men - Linear Effects Over Time

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-54)
exposure	0.047	0.059*	0.035	0.087**	−0.032
	(0.032)	(0.031)	(0.041)	(0.042)	(0.041)
Num.Obs.	30 572	10 764	8572	6860	4376
R2 Adj.	0.080	0.079	0.087	0.081	0.086
Mean Y	0.816	0.816	0.816	0.818	0.811
SD Y	0.285	0.284	0.282	0.287	0.292
Controls	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓

Panel D: Men - Quadratic Effects Over Time

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-54)
exposure2	0.039	0.056	0.022	0.079**	−0.037
	(0.032)	(0.034)	(0.043)	(0.036)	(0.043)
Num.Obs.	30 572	10 764	8572	6860	4376
R2 Adj.	0.080	0.079	0.086	0.080	0.086
Mean Y	0.816	0.816	0.816	0.818	0.811
SD Y	0.285	0.284	0.282	0.287	0.292
Controls	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓

Table 7: Autonomy of women (share) with respect to HH decision-making and mobility and men's beliefs towards the share of decisions women should participate in

Note: The tables show separate regressions of autonomy by age cohort on exposure to radio. For women, autonomy is defined as the share of decisions a woman participates in / places she can visit on her own. For men, the variable is defined as the share of decisions he believes a woman should participate in. Panels A and C assume linear effects over time and Panel B and D quadratic effects. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

from 2016-20. In total, 84 radios launched post-2015 are included in the placebo data, covering 6,620 survey clusters (in comparison to 8k in the main data). Interestingly, 51% of observations in the placebo sample are also part of the main sample given their vicinity to both a station that launches pre- and post-2015. This suggests comparable radio placement patterns between the two periods as also visible in Figure B.5 (in the appendix). Tables F.3 and F.4 (in the appendix) show the results of all main regressions for the placebo sample allowing for linear or quadratic effects over time. The regressions show no effects on outcomes. This speaks against heterogeneity in unobservables being correlated with the variation in radio exposure used in this paper.

A final threat to identification is the possibility of a change happening in coverage areas simultaneously to but independently of the launch of radios. To interfere with the variation I am using, such a change would have to closely follow radios' coverage areas rather than, e.g., being related to the travel time between survey clusters and the radio tower for which I control. In addition, it would have to be closely associated with the timing of radio stations' launch date. Although such a change is difficult to imagine, one piece of evidence speaks against it. Specifically, thinking of education, such a change could be a supply side effect of schooling (though, again, it seems unlikely that such an effect would closely follow the radio stations' coverage area). In Tables D.3 and D.4 (in the appendix), I study changes in reasons for school dropout. If such an effect were present, effects would likely be driven by the availability of a school for girls or improved transport to school. I find no evidence for such supply side factors speaking against this being a driver of the results. While this speaks against this type of endogeneity, it is further important to acknowledge that community radio may, of course, drive such change. For instance, if listeners put more value on education, this may increase demand and, ultimately, supply thereof (e.g., through economic or political channels).

Finally, I test the robustness of the findings with respect to regression specifications. The above results show that the results are robust to varying the functional form of the treatment variable or definitions of the dependent variable (in the context of educational outcomes). In addition, Section G (in the appendix) varies the regression framework by applying different standard errors, clustered at the subdistrict level) and varying the distance threshold of data inclusion. The results are robust to any of the above changes. Finally, I also show that the results are robust to computing the coverage probability. While I explicitly compute the probability density function of potential survey locations for the paper's main results, the results replicate when simply assuming a uniform distribution (see Table E.2 and E.3 in appendix).

Overall, the robustness and placebo checks support the causal interpretation of the effects of the treatment variation exploited in this paper.

VI.5 Evaluating the Jittering Correction

Finally, I compare the results in the main regressions presented in Sections VI.2 and VI.3 to those if I had not corrected for jittering. For results without the correction, I simply measure whether the location as reported by the NFHS lies within the treatment area. To get a measure of exposure, I multiply the dummy variable by the share of time the respective radio has been present in the region. The variable is equal to the exposure variable for locations certainly covered or not covered by the radio signal. It only differs for location in the vicinity of the coverage area. In addition, I change distance controls to simply control for the line of sight between the reported location and the radio tower (instead of the expected distance).

The results on all main outcomes are presented in Table E.4 (in the appendix). These show that correcting for the jittering substantially improves the precision of estimates and suggest significant improvements due to a reduction in attenuation bias. On average, coefficient sizes increase by 65% when correcting for the jittering. This is in line with a substantial reduction of attenuation bias, which would downward bias coefficients due to measurement error in the treatment variable. It may also be explained by mismeasurement of the distance variable. Specifically, Table E.5 in the appendix shows that simply correcting for the distance substantially affects estimates.

The above results suggest that the correction that I propose substantially improves estimates when working with jittered survey data. This is likely particularly relevant in settings with scattered treatment or coverage areas as well as when studying phenomena that are relatively local when compared to the distance across which the jittering is performed. Overall, it suggests that my approach opens the path for study designs rendered infeasible using current approaches to correct for the jittering. It more generally allows researchers to obtain consistent estimate in light of jittered survey data.

VII Summary and Concluding Remarks

This paper provides evidence on the long-term and large-scale use of grassroots media as a policy instrument. For this, I evaluate a policy India enacted in 2006, which grants educational institutions and NGOs radio licenses with the requirement to focus on local development issues. Based on information gathered on the content of community radio stations, I identify women's empowerment as a key theme in radio programming. To identify the effects of radio, I rely upon a standard approach that exploits topographic features between radio towers and receivers. I further combine this with a novel econometric method to uncover parameters in light of jittered survey locations. The results show that community radio stations have substantial effects on attitudes and behavior of and toward women and girls. Areas exogenously exposed to community radio show increased education and degree completion rates for girls. Ancillary results suggest that changes in parents' aspirations explain these results. In line with this, young women marry later and have fewer kids. I find particularly strong decreases in child marriage rates. I also present evidence for greater autonomy of young women and of men being more supportive of women's autonomy. Suggestive results further point toward changes in women's attitudes toward domestic violence and fewer experiences thereof, suggesting no 'male backlash' in response to women's empowerment.

Overall, the results demonstrate that grassroots media can be used as a large-scale and long-term policy instrument to affect development outcomes. These insights complement and go beyond earlier research which largely focuses on the unintended impacts of entertainment media or experiments (DellaVigna and La Ferrara, 2015) as well as findings on single issue government campaigns (Khalifa, 2022). Grassroots media policies akin India's may serve as an effective policy tool for developing countries. Given limited government resources, the policy provides a way to draw on 'civil society's' (i.e. NGOs' and educational institutions') resources and knowledge to affect development outcomes. Local institutions' knowledge of local issues is likely to be particularly valuable in culturally and linguistically diverse countries, a characteristic India shares with much of the developing world. Further, community radio can potentially address populations with little trust in the government and, hence, government media campaigns.

While radio remains an integral part of most countries' media spheres, an important question for future research and policy making is how the concept of community radio can be translated into other types of media. Some community radio stations have already taken first steps, e.g., by joining social media or broadcasting online.[26] In addition, research on other themes of community radio programming would be an important addition to this paper's insights. Although this paper focuses on women's empowerment and education-related outcomes, the content analyses suggest that radios discuss a variety of other topics. For instance, future research may evaluate effects on agricultural yields, health or the uptake of government schemes. The results further speak to India's policy in particular. While India is very diverse and inhabits 17% of the world's population, the policy may function differentially in other contexts (U.S. Census Bureau, August 2024). It would therefore be important to expand the evidence to other countries. South Asia may be a good place to start, as countries, like Bangladesh, passed similar community radio policies at around the same time as India (Raghunath, 2020). Another interesting avenue for future research would be a closer investigation of the channels driving effects of community radio. While I provide evidence on potential mechanisms driving the results, field work, such as through RCTs, may be a viable path to gather more precise insights. Finally, in addition to the topical contribution, my paper also suggests a novel approach

[26] For example, `radio.garden` features many community radio stations around the world, including India.

to deal with spatially jittered survey data. As I demonstrate, the correction strongly improves on attenuation bias. This opens the path for future research using such data, especially when working in settings where the jittering imposes challenges to identify effects and potentially deems previous research designs infeasible. It also, more generally, allows researchers to uncover true parameters in light of the jittering. Such research would also help to better understand under which circumstances the approach yields the largest benefits and where its application is less beneficial, e.g., because treatment areas are sufficiently large or do not matter as much for identification.

References

Adena, Maja, Ruben Enikolopov, Maria Petrova, and Hans-Joachim Voth, "Bombs, Broadcasts and Resistance: Allied Intervention and Domestic Opposition to the Nazi Regime During World War II," *SSRN Electronic Journal*, 2020.

— , — , — , Veronica Santarosa, and Ekaterina Zhuravskaya, "Radio and the Rise of The Nazis in Prewar Germany," *The Quarterly Journal of Economics*, November 2015, *130(4)* (4), 1885–1939.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn, "On the Origins of Gender Roles: Women and the Plough," *The Quarterly Journal of Economics*, May 2013, *128* (2), 469–530.

Altay, Umut, John Paige, Andrea Riebler, and Geir-Arne Fuglstad, "Fast Geostatistical Inference under Positional Uncertainty: Analysing DHS Household Survey Data," December 2022.

— , — , — , and — , "Jittering Impacts Raster- and Distance-based Geostatistical Analyses of DHS Data," November 2022.

Anderson, Jason and Amy Lightfoot, "The School Education System in India: An Overview," Technical Report, British Council, New Delhi July 2019.

Arel-Bundock, Vincent, Joachim Gassen, Nathan Eastwood, Nick Huntington-Klein, Moritz Schwarz, Benjamin Elbers (0000-0001-5392-3448), Grant McDermott, and Lukas Wallrich, "Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready," October 2023.

Arias, Eric, "Media, Common Knowledge, and Violence Against Women: A Field Experiment on Norms Change in Mexico," Working Paper November 2014.

Armand, Alex, Paul Atwell, and Joseph F. Gomes, "The Reach of Radio: Ending Civil Conflict through Rebel Demobilization," *American Economic Review*, May 2020, *110* (5), 1395–1429.

Asher, Sam, Paul Novosad, and Charlie Rafkin, "Intergenerational Mobility in India: New Methods and Estimates Across Time, Space, and Communities," February 2021.

Baird, Sarah, Craig McIntosh, and Berk Özler, "Cash or Condition? Evidence from a Cash Transfer Experiment *," *The Quarterly Journal of Economics*, November 2011, *126* (4), 1709–1753.

Balk, D., M. R. Montgomery, H. Engin, E. Major, and B. Jones, "Spatial Data from the 2011 India Census," 2020.

Balk, Deborah, Mark R. Montgomery, Hasim Engin, Natalie Lin, Elizabeth Major, and Bryan Jones, "Urbanization in India: Population and Urban Classification Grids for 2011," *Data*, February 2019, *4* (1), 35.

Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman, "Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa," *American Economic Journal: Applied Economics*, January 2020, *12* (1), 210–259.

Banerjee, A. V., Eliana La Ferrara, and Victor Orozco, "The Entertaining Way to Behavioral Change: Fighting HIV with MTV," Working Paper 2019.

Banerjee, Abhijit, Eliana La Ferrara, and Victor Orozco, "Entertainment, Education, and Attitudes Toward Domestic Violence," *AEA Papers and Proceedings*, May 2019, *109*, 133–137.

Basu, Alaka Malwade, "Why Does Education Lead to Lower Fertility? A Critical Review of Some of the Possibilities," *World Development*, October 2002, *30* (10), 1779–1790.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo, "Quanteda: An R Package for the Quantitative Analysis of Textual Data," *Journal of Open Source Software*, October 2018, *3* (30), 774.

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*, October 2017, *15* (5), 1025–1055.

Bergé, Laurent, "Efficient Estimation of Maximum Likelihood Models with Multiple Fixed-Effects: The R Package FENmlm," *DEM Discussion Paper Series*, 2018, (18-13).

Bernard, Tanguy, Stefan Dercon, Kate Orkin, and Alemayehu Seyoum Taffesse, "The Future in Mind: Aspirations and Forward-Looking Behaviour in Rural Ethiopia," *CEPR Discussion Paper*, 2014, (DP10224).

Biswas, Bapan and Nasrin Banu, "Economic Empowerment of Rural and Urban Women in India: A Comparative Analysis," *Spatial Information Research*, February 2023, *31* (1), 73–89.

Bjorvatn, Kjetil, Alexander W. Cappelen, Linda Helgesson Sekei, Erik Ø. Sørensen, and Bertil Tungodden, "Teaching Through Television: Experimental Evidence on Entrepreneurship Education in Tanzania," *Management Science*, June 2020, *66* (6), 2308–2325.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 2003, *3* (Jan), 993–1022.

Boas, Taylor C. and F. Daniel Hidalgo, "Controlling the Airwaves: Incumbency Advantage and Community Radio in Brazil," *American Journal of Political Science*, 2011, *55* (4), 869–885.

Burgert, Clara R., Josh Colston, Roy Thea, and Blake Zachary, "Geographic Displacement Procedure and Georeferenced Data Release: Policy for the Demographic and Health Surveys," Technical Report 7, ICF International, United States Agency for International Development (USAID), Calverton, Maryland September 2013.

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott, "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, October 2020, *110* (10), 2997–3029.

_ , Alexander W. Cappelen, Bertil Tungodden, Alessandra Voena, and David H. Yanagizawa-Drott, "How Are Gender Norms Perceived?," March 2023.

_ and Davide Cantoni, "A Tear in the Iron Curtain: The Impact of Western Television on Consumption Behavior," *Review of Economics and Statistics*, March 2016, *98* (1), 25–41.

Cassidy, Rachel, Anaya Dam, Wendy Janssens, Umair Kiani, and Karlijn Morsink, "Father of the Bride, or Steel Magnolias? Targeting Men, Women or Both to Reduce Child Marriage," Working Paper 22/50, lnstitute for Fiscal Studies November 2022.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng, "On Binscatter," *American Economic Review*, 2024, *5* (114).

Census of India, "Census of India : General Note," https://web.archive.org/web/20220314142033/https://www.censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/gen_note.html 2002.

Chang, Wei, Lucia Diaz-Martin, Akshara Gopalan, Eleonora Guarnieri, Seema Jayachandran, and Claire Walsh, "What Works to Enhance Women's Agency: Cross-cutting Lessons from Experimental and Quasi-Experimental Studies," *J-PAL Working Paper*, July 2020, *87.*

Chong, Alberto and Eliana La Ferrara, "Television and Divorce: Evidence from Brazilian Novelas," *Journal of the European Economic Association*, May 2009, *7* (2-3), 458–468.

Conley, T. G., "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, September 1999, *92* (1), 1–45.

Conley, Timothy G., "Spatial Econometrics," in Steven N. Durlauf and Lawrence E. Blume, eds., *Microeconometrics*, The New Palgrave Economics Collection, London: Palgrave Macmillan UK, 2010, pp. 303–313.

Coville, Aidan, Vincenzo Di Maro, Felipe Alexander Dunsch, and Siegfried Zottel, "The Nollywood Nudge: An Entertaining Approach to Saving," Policy Research Working Paper 8920, The World Bank, Washington D.C. June 2019.

CRFC, "Searchable Database of CR Stations," http://crfc.in/list-of-cr-station/?status=operational-station 2021.

_ , "About DAVP," http://crfc.in/about-davp/ 2022.

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

_ , Ruben Enikolopov, Vera Mironova, Maria Petrova, and Ekaterina Zhuravskaya, "Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia," *American Economic Journal: Applied Economics*, July 2014, *6* (3), 103–132.

DHS, "Guide to DHS Statistics DHS-8: Analyzing DHS Data," https://dhsprogram.com/data/Guide-to-DHS-Statistics/Analyzing_DHS_Data.htm.

Doepke, Matthias, Michèle Tertilt, and Alessandra Voena, "The Economics and Politics of Women's Rights," *Annual Review of Economics*, September 2012, *4* (Volume 4, 2012), 339–372.

Duflo, Esther, "Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa," *The World Bank Economic Review*, June 2003, *17* (1), 1–25.

_ , "Women Empowerment and Economic Development," *Journal of Economic Literature*, December 2012, *50* (4), 1051–1079.

_ , Pascaline Dupas, and Michael Kremer, "Education, HIV, and Early Fertility: Experimental Evidence from Kenya," *American Economic Review*, September 2015, *105* (9), 2757–2797.

Durante, Ruben, Paolo Pinotti, and Andrea Tesei, "The Political Legacy of Entertainment TV," *American Economic Review*, July 2019, *109* (7), 2497–2530.

Enikolopov, Ruben and Maria Petrova, "Mass Media and Its Influence on Behaviour," Els Opuscles Del CREI 44, Centre de Recerca en Economia Internacional, Barcelona, Spain 2017.

_ , _ , and Ekaterina Zhuravskaya, "Media and Political Persuasion: Evidence from Russia," *The American Economic Review*, 2011, *101* (7), 3253–3285.

Ferrara, Eliana La, "Mass Media and Social Change: Can We Use Television to Fight Poverty?," *Journal of the European Economic Association*, August 2016, *14* (4), 791–827.

Fraser, Colin and Sonia Restrepo Estrada, "Community Radio Handbook," Handbook, UNESCO, Paris 2001.

_ and Sonia Restrepo-Estrada, "Community Radio for Change and Development," *Development*, December 2002, *45* (4), 69–73.

Giacobino, Hélène, Elise Huillery, Bastien Michel, and Mathilde Sage, "Schoolgirls, Not Brides: Education as a Shield against Child Marriage," *American Economic Journal: Applied Economics*, October 2024, *16* (4), 109–143.

Gius, Evelyn, Jan Christoph Meister, Malte Meister, Marco Petris, Mareike Schumacher, and Dominik Gerstorfer, "CATMA," Zenodo May 2023.

Glennerster, R., J. Murray, and V. Pouliquen, "The Media or the Message? Experimental Evidence on Mass Media and Modern Contraception Uptake in Burkina Faso," 2021.

Goldin, Claudia, "The Quiet Revolution That Transformed Women's Employment, Education, and Family," *American Economic Review*, May 2006, *96* (2), 1–21.

Govt. of India, "Policy Guidelines for Setting up Community Radio Stations in India," Policy Guidelines, Government of India, New Delhi December 2006.

Green, Donald P. and Srinivasan Vasudevan, "Diminishing the Effectiveness of Vote Buying: Experimental Evidence from a Persuasive Radio Campaign in India," Working Paper 2018.

_ , Anna Wilke, and Jasper Cooper, "Silence Begets Violence: A Mass Media Experiment to Prevent Violence against Women in Rural Uganda," *Newhaven, CT: Innovations for Poverty*, 2018.

Grün, Bettina and Kurt Hornik, "Topicmodels: An R Package for Fitting Topic Models," *Journal of Statistical Software*, May 2011, *40*, 1–30.

Guarnieri, Eleonora and Helmut Rainer, "Colonialism and Female Empowerment: A Two-Sided Legacy," *Journal of Development Economics*, June 2021, *151*, 102666.

Hansen, Stephen, Michael McMahon, and Andrea Prat, "Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach*," *The Quarterly Journal of Economics*, May 2018, *133* (2), 801–870.

Heath, Rachel, Arielle Bernhardt, Girija Borker, Anne Fitzpatrick, Anthony Keats, Madeline McKelway, Andreas Menzel, Teresa Molina, and Garima Sharma, "Female Labour Force Participation," *VoxDev Lit*, February 2024, *11* (1).

Hu, Yingyao, "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution," *Journal of Econometrics*, May 2008, *144* (1), 27–61.

Hufford, George A., "The ITS Irregular Terrain Model, Version 1.2. 2, the Algorithm," Technical Report 2002.

IIPS and ICF, *National Family Health Survey (NFHS-4), 2015-16: India.*, Mumbai: International Institute for Population Sciences (IIPS), 2017.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, B. Puranen et al., "World Values Survey: Round Five - Country-Pooled Datafile Version," http://www.worldvaluessurvey.org/WVSDocumentationWV5.jsp 2014.

Jacob, Jose, "Community Radio Stations in India," https://qsl.net/vu2jos/fm/cr.htm 2021.

Jarvis, A., A. Nelson, and E. Guevara, "Hole-Filled Seamless SRTM Data V4," Data Set, International Centre for Tropical Agriculture (CIAT), 2008.

Jayachandran, Seema, "The Roots of Gender Inequality in Developing Countries," *Annual Review of Economics*, 2015, *7* (1), 63–88.

_ , "Odds Are You're Measuring Son Preference Incorrectly," https://blogs.worldbank.org/impactevaluations/odds-are-you-re-measuring-son-preference-incorrectly 2017.

_ **and Adriana Lleras-Muney**, "Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines*," *The Quarterly Journal of Economics*, February 2009, *124* (1), 349–397.

Jensen, Robert and Emily Oster, "The Power of TV: Cable Television and Women's Status in India*," *The Quarterly Journal of Economics*, August 2009, *124* (3), 1057–1094.

Karra, Mahesh, David Canning, and Ryoko Sato, "Adding Measurement Error to Location Data to Protect Subject Confidentiality While Allowing for Consistent Estimation of Exposure Effects," *Journal of the Royal Statistical Society Series C: Applied Statistics*, November 2020, *69* (5), 1251–1268.

Kasampalis, Stylianos, Pavlos I. Lazaridis, Zaharias D. Zaharis, Aristotelis Bizopoulos, Spiridon Zettas, and John Cosmas, "Comparison of Longley-Rice, ITM and ITWOM Propagation Models for DTV and FM Broadcasting," in "2013 16th International Symposium on Wireless Personal Multimedia Communications (WPMC)" June 2013, pp. 1–6.

Kasteng, Frida, Joanna Murray, Simon Cousens, Sophie Sarrassat, Jennifer Steel, Nicolas Meda, Moctar Ouedraogo, Roy Head, and Josephine Borghi, "Cost-Effectiveness and Economies of Scale of a Mass Radio Campaign to Promote Household Life-Saving Practices in Burkina Faso," *BMJ Global Health*, July 2018, *3* (4), e000809.

Kearney, Melissa S. and Phillip B. Levine, "Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing," *American Economic Review*, December 2015, *105* (12), 3597–3632.

Keats, Anthony, "Women's Schooling, Fertility, and Child Health Outcomes: Evidence from Uganda's Free Primary Education Program," *Journal of Development Economics*, November 2018, *135*, 142–159.

Keefer, Philip and Stuti Khemani, "The Government Response to Informed Citizens : New Evidence on Media Access and the Distribution of Public Health Benefits in Africa," *World Bank Economic Review*, 2016.

Khalifa, Suzanna, "Female Genital Cutting and Bride Price," Working Paper December 2022.

KPMG, "Media for the Masses: The Promise Unfolds," India Media and Entertainment Industry Report 2017.

Kumar, Kanchan, "Mixed Signals: Radio Broadcasting Policy in India," *Economic and Political Weekly*, 2003, *38* (22), 2173–2182.

Lewbel, Arthur, "Estimation of Average Treatment Effects with Misclassification," *Econometrica*, 2007, *75* (2), 537–551.

Lowes, Sara, "Culture in Historical Political Economy," September 2022.

Lucas, Adrienne M. and Isaac M. Mbiti, "Does Free Primary Education Narrow Gender Differences in Schooling? Evidence from Kenya†," *Journal of African Economies*, November 2012, *21* (5), 691–722.

Maertens, Annemie, "Social Norms and Aspirations: Age of Marriage and Education in Rural India," *World Development*, July 2013, *47*, 1–15.

Mahajan, Aprajit, "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 2006, *74* (3), 631–665.

Michler, Jeffrey D., Anna Josephson, Talip Kilic, and Siobhan Murray, "Privacy Protection, Measurement Error, and the Integration of Remote Sensing and Socioeconomic Survey Data," *Journal of Development Economics*, September 2022, *158*, 102927.

Miller, Grant, "Contraception as Development? New Evidence from Family Planning in Colombia," *The Economic Journal*, June 2010, *120* (545), 709–736.

MOIB, "State Wise Details of Operational Community Radio Stations in India as on 31-03-2020," List of Operational Community Radio Stations, Ministry of Information and Broadcasting, New Delhi March 2020.

Mookerjee, Sulagna, "Gender-Neutral Inheritance Laws, Family Structure, and Women's Status in India," *The World Bank Economic Review*, June 2019, *33* (2), 498–515.

Morgan-Wall, Tyler, "Rayshader: Create Maps and Visualize Data in 2D and 3D," February 2024.

Murray, Joanna, Pieter Remes, Rita Ilboudo, Mireille Belem, Souleymane Salouka, Will Snell, Cathryn Wood, Matthew Lavoie, Laurent Deboise, and Roy Head, "The Saturation+ Approach to Behavior Change: Case Study of a Child Survival Radio Campaign in Burkina Faso," *Global Health: Science and Practice*, November 2015, *3* (4), 544–556.

Myers, Mary, "Voices from Villages: Community Radio in the Developing World," CIMA Research Report, Center for International Media Assistance, Washington D.C. 2011.

Nirmala, Yalala, "The Role of Community Radio in Empowering Women in India," *Media Asia*, April 2015, *42* (1-2), 41–46.

Okuyama, Yoko, "Empowering Women Through Radio: Evidence from Occupied Japan," 2023.

Olken, Benjamin A., "Do Television and Radio Destroy Social Capital? Evidence from Indonesian Villages," *American Economic Journal: Applied Economics*, October 2009, *1* (4), 1–33.

Pavarala, Vinod and Kanchan K. Malik, *Other Voices: The Struggle for Community Radio in India*, New Delhi: Sage Publications India, 2007.

Pebesma, Edzer, "Simple Features for R: Standardized Support for Spatial Vector Data," *The R Journal*, 2018, *10* (1), 439–446.

Pennec, Erwan Le and Kamil Slowikowski, "Ggwordcloud: A Word Cloud Geom for 'Ggplot2'," 2023.

Qian, Ruoyu, "Communist Propaganda and Women's Status," *Journal of Development Economics*, October 2024, *171*, 103341.

Raghunath, Preeti, "A Glocal Public Sphere: Opening up of Radio to Communities in South Asia," in "Community Radio Policies in South Asia: A Deliberative Policy Ecology Approach," Singapore: Springer Singapore, 2020, pp. 145–192.

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*, December 2015, *132*, 13–22.

Riley, Emma, "Role Models in Movies: The Impact of Queen of Katwe on Students' Educational Attainment," *The Review of Economics and Statistics*, March 2024, *106* (2), 334–351.

Rusche, Felix, "Chapter 17: Large Language Model APIs," in Paul C. Bauer, Camille Landesvatter, and Lion Behrens, eds., *APIs for Social Scientists: A Collaborative Review*, 2024.

Schennach, Susanne M., "Recent Advances in the Measurement Error Literature," *Annual Review of Economics*, October 2016, *8* (Volume 8, 2016), 341–377.

Sen, Amartya, "Missing Women.," *BMJ: British Medical Journal*, 1992, *304* (6827), 587.

SMART, "Gender Analyses in Community Radio Stations, 2023," Technical Report, SMART NGO, New Delhi March 2023.

Thomas, Pradio Ninan, "Retuning All India Radio: We Need to Collectively Reimagine All India Radio as an Independent Public Service Broadcaster," *Economic and Political Weekly*, 2013, *48* (40), 83–83.

UNESCO, "Statistics on Radio," 2013.

U.S. Census Bureau, "U.S. and World Population Clock," Technical Report, U.S. Census Bureau August 2024.

Varshney, Deepak, Pramod K. Joshi, Anjani Kumar, Ashok K. Mishra, and Shantanu Kumar Dubey, "Examining the Transfer of Knowledge and Training to Smallholders in India: Direct and Spillover Effects of Agricultural Advisory Services in an Emerging Economy," *World Development*, December 2022, *160*, 106067.

Walsh, Christoph, "Social Impacts of New Radio Markets in Ghana: A Dynamic Structural Analysis," 2023.

Watanabe, Kohei and Phan Xuan-Hieu (GibbsLDA++), "Seededlda: Seeded Sequential LDA for Topic Modeling," May 2023.

Weiss, D., Andy Nelson, H. S. Gibson, W. Temperley, Stephen Peedell, A. Lieber, M. Hancher, Eduardo Poyart, Simão Belchior, and Nancy Fullman, "A Global Map of Travel Time to Cities to Assess Inequalities in Accessibility in 2015," *Nature*, 2018, *553* (7688), 333–336.

Wickham, Hadley, "Ggplot2," *WIREs Computational Statistics*, March 2011, *3* (2), 180–185.

World Bank, "Literacy Rate, Adult Total (% of People Ages 15 and above) - India," https://-data.worldbank.org 2023.

WorldPop, "The Spatial Distribution of Population in 2020, India," 2020.

Yanagizawa-Drott, David, "Propaganda and Conflict: Evidence from the Rwandan Genocide," *The Quarterly Journal of Economics*, November 2014, *129* (4), 1947–1994.

Online Appendix for

Broadcasting Change: India's Community Radio Policy and Women's Empowerment

by

Felix Rusche

Table of Contents

A Spatial Jittering: Expected Distance

The Elliptic Integral of the Second Kind can be expressed as follows:

$$EI(x) = \int_0^{\frac{\pi}{2}} \sqrt{1 - x^2 \sin^2(\phi)} d\phi \tag{18}$$

The distance formula can be re-formulated as the Elliptic Integral of Second Kind:

$$\begin{aligned}
\mathbb{E}(d(x^*, t)|r) &= \frac{1}{2\pi} \int_0^{2\pi} \sqrt{r^2 + d(x,t)^2 - 2rd(x,t)\cos(\phi)} d\phi && \mid \text{ By symmetry of circle} \\
&= \frac{1}{\pi} \int_0^{\pi} \sqrt{r^2 + d(x,t)^2 + 2rd(x,t)\cos(\phi)} d\phi && \mid \text{ Define: } \phi = 2k \\
&= \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \sqrt{r^2 + d(x,t)^2 + 2rd(x,t)\cos(2k)} dk && \mid \cos(2k) = 1 - 2\sin^2(k) \\
&= \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \sqrt{r^2 + d(x,t)^2 + 2rd(x,t)(1 - 2\sin^2(k))} dk \\
&= \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \sqrt{(r + d(x,t))^2 - 4rd(x,t)\sin^2(k)} dk \\
&= \frac{2}{\pi}(r + d(x,t)) \int_0^{\frac{\pi}{2}} \sqrt{1 - \frac{4rd(x,t)}{(r + d(x,t))^2} \sin^2(k)} dk \\
&= \frac{2}{\pi}(r + d(x,t)) \int_0^{\frac{\pi}{2}} \sqrt{1 - (\frac{2\sqrt{rd(x,t)}}{r + d(x,t)})^2 \sin^2(k)} dk && \mid \text{ Following Eq. 18} \\
&= \frac{2}{\pi}(r + d(x,t)) EI(\frac{2\sqrt{rd(x,t)}}{r + d(x,t)})
\end{aligned} \tag{19}$$

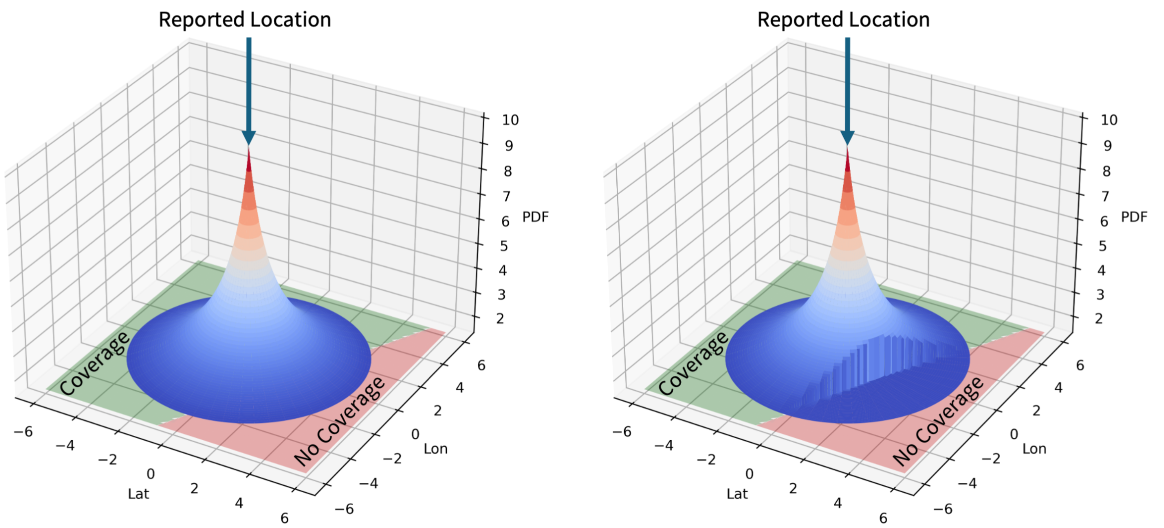


Figure A.1: Visualization of jittering correction

Note: The figures above show the PDF of x^* conditional on observing x (reported location) for the simplified case. The figure on the left shows the full likelihood mass, and the one on the right the likelihood mass on the treatment area. The PDF is rescaled for illustrative purposes. While it follows a similar circus tent shape, it is 'steeper' in reality.

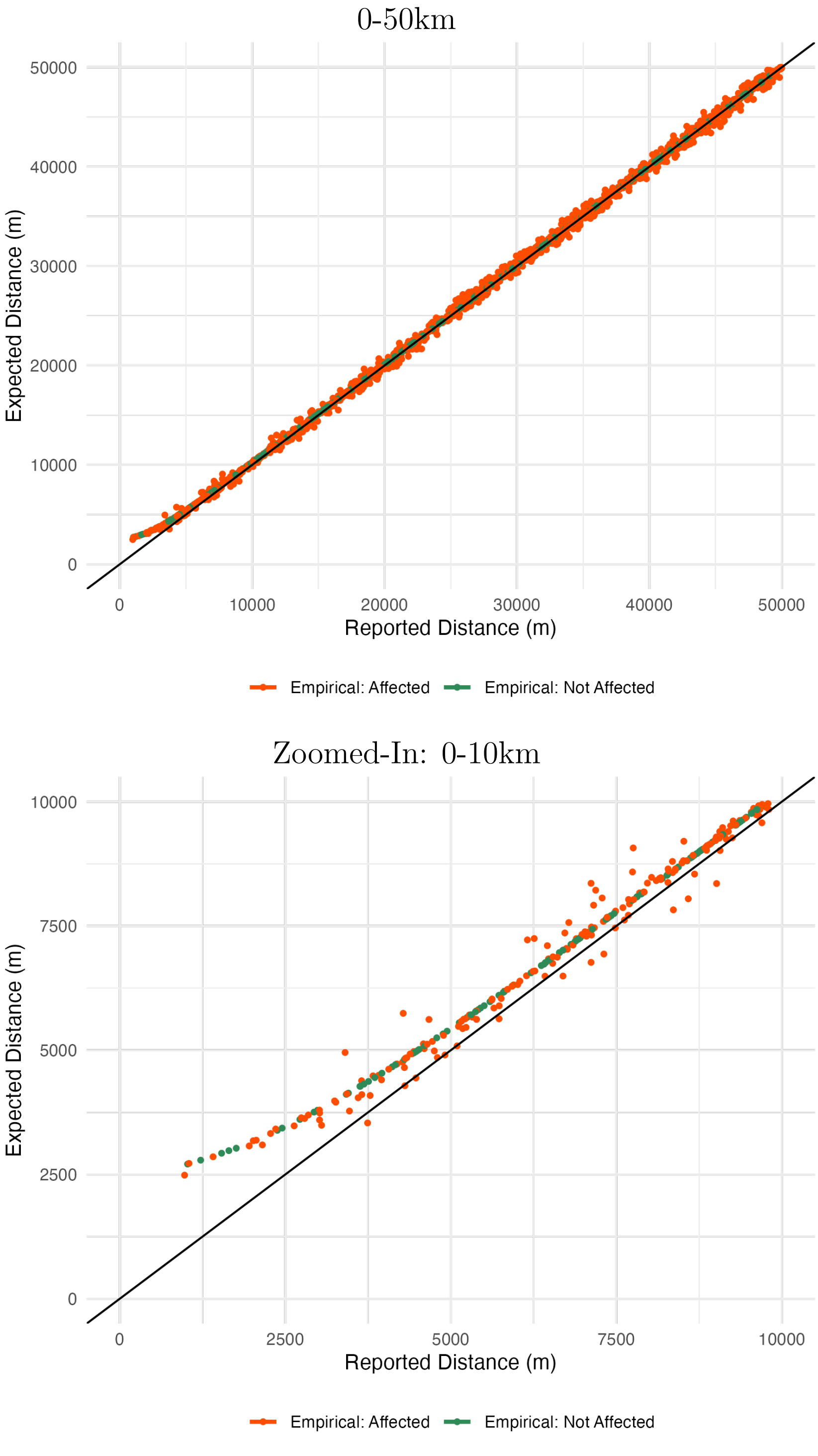


Figure A.2: Comparison: reported and expected difference

Note: The above graphs compare three different distance measures: The x-axis shows the distance between the closest radio tower and a given DHS cluster as computed based on the displaced location indicated in the DHS data. The y-axis provides the Expected Distance between the radio and the DHS observation taking into account the displacement. The orange line ("Simulation") compares the reported and expected distance based upon Equation 7. The dots ("Empirical") indicate the expected distance as simulated using a grid around reported locations. These are split into two groups: the unaffected group includes locations whose displacement was not affected by a district border. For these, the results should hold as reported in Equation 7. For the affected group, the expected distance can vary from the equation due to the district border. The results show that this is indeed the case. While the unaffected locations lie on the simulation line, the affected ones vary slightly from it. Further, it is visible that the displacement mainly affects distances within the first 10km.

B Descriptives

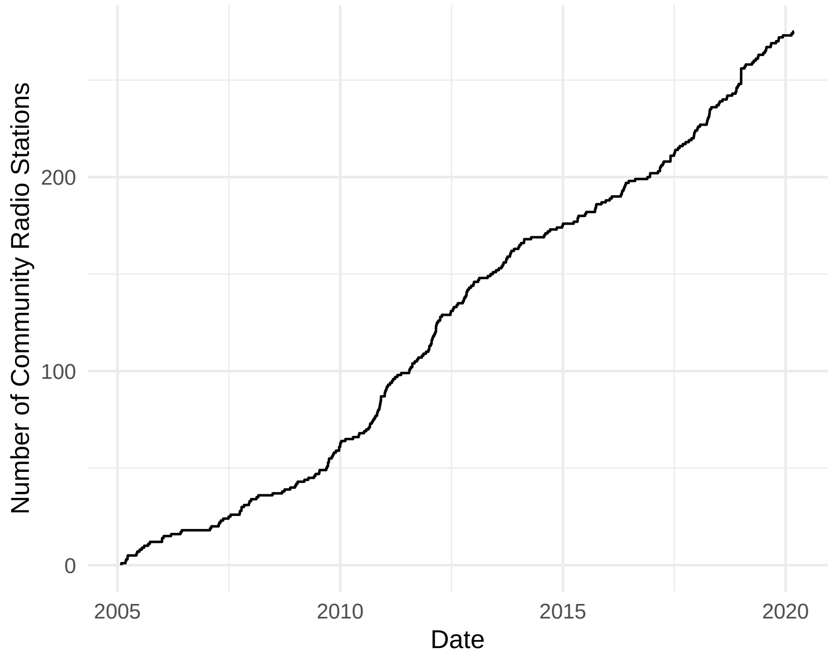


Figure B.1: Total number of community radio stations on air by date

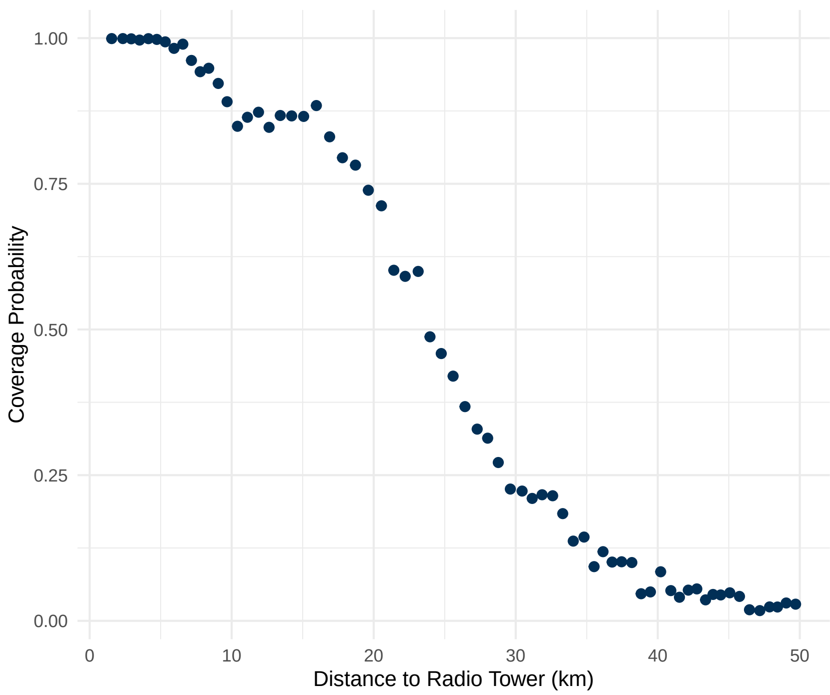


Figure B.2: Binscatter plot of coverage probability and distance to radio

Note: The plot is created based on the *binsreg* package in R (Cattaneo et al., 2024). The number of bins is determined by the IMSE-optimal direct plug-in rule. The underlying data are the pooled coverage probabilities and distances to the first radio station from both the main and placebo data.

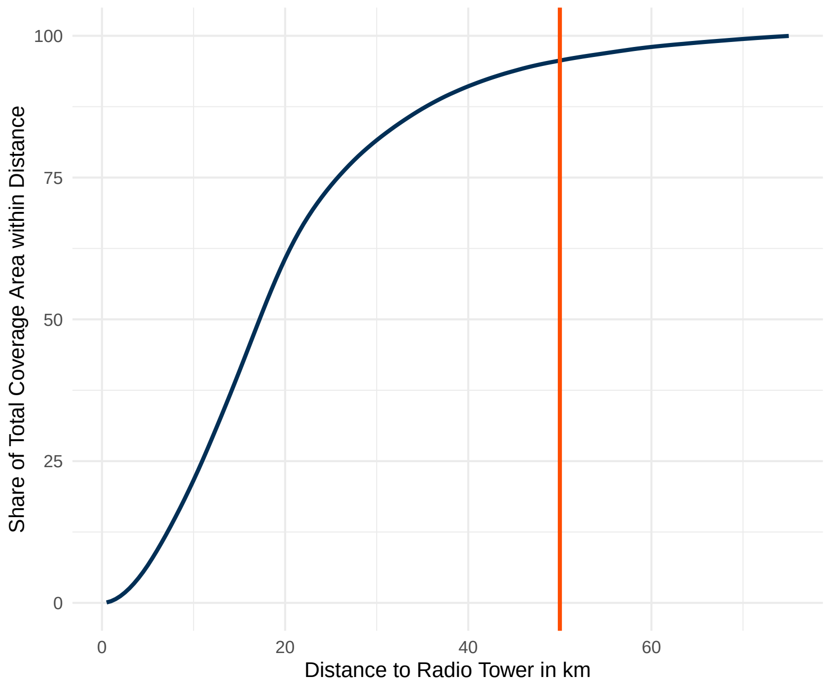


Figure B.3: Share of total coverage area within distance

Note: The above graph visualizes the share of CRs' total coverage area by distance to the radio tower. 58% lies within 20km, 81% within 30km, 91% within 40km, 96% within 50km, 98% within 60km, and 99.9% within 75km.

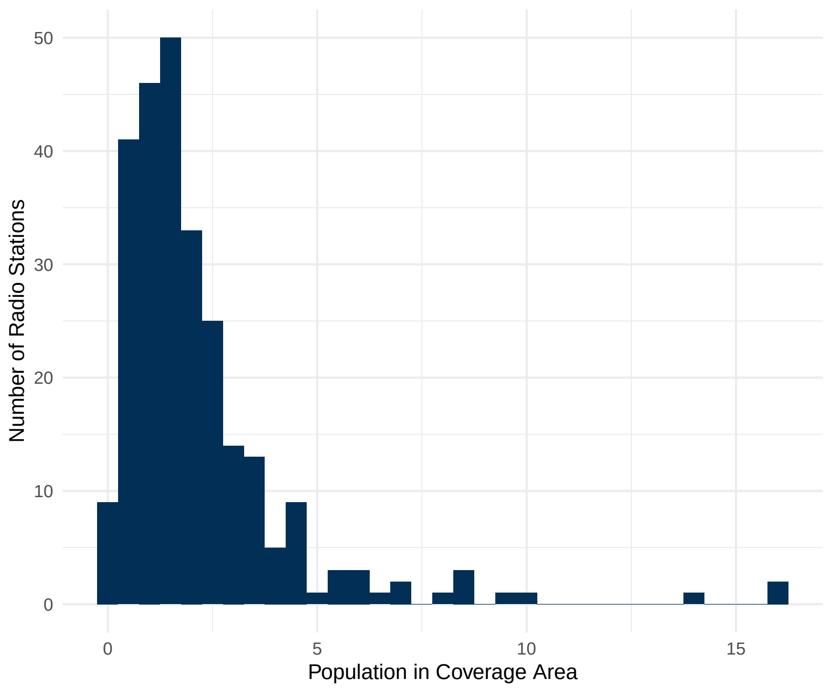


Figure B.4: Estimated population within coverage area

Note: The above figure includes information on the total population within reach of 264 radio stations. 2016 Population estimates are based on WorldPop (2020). Note that the total number of individuals reached by any community radio is not equal to the sum of the population reached by the radios above, given that coverage areas overlap.

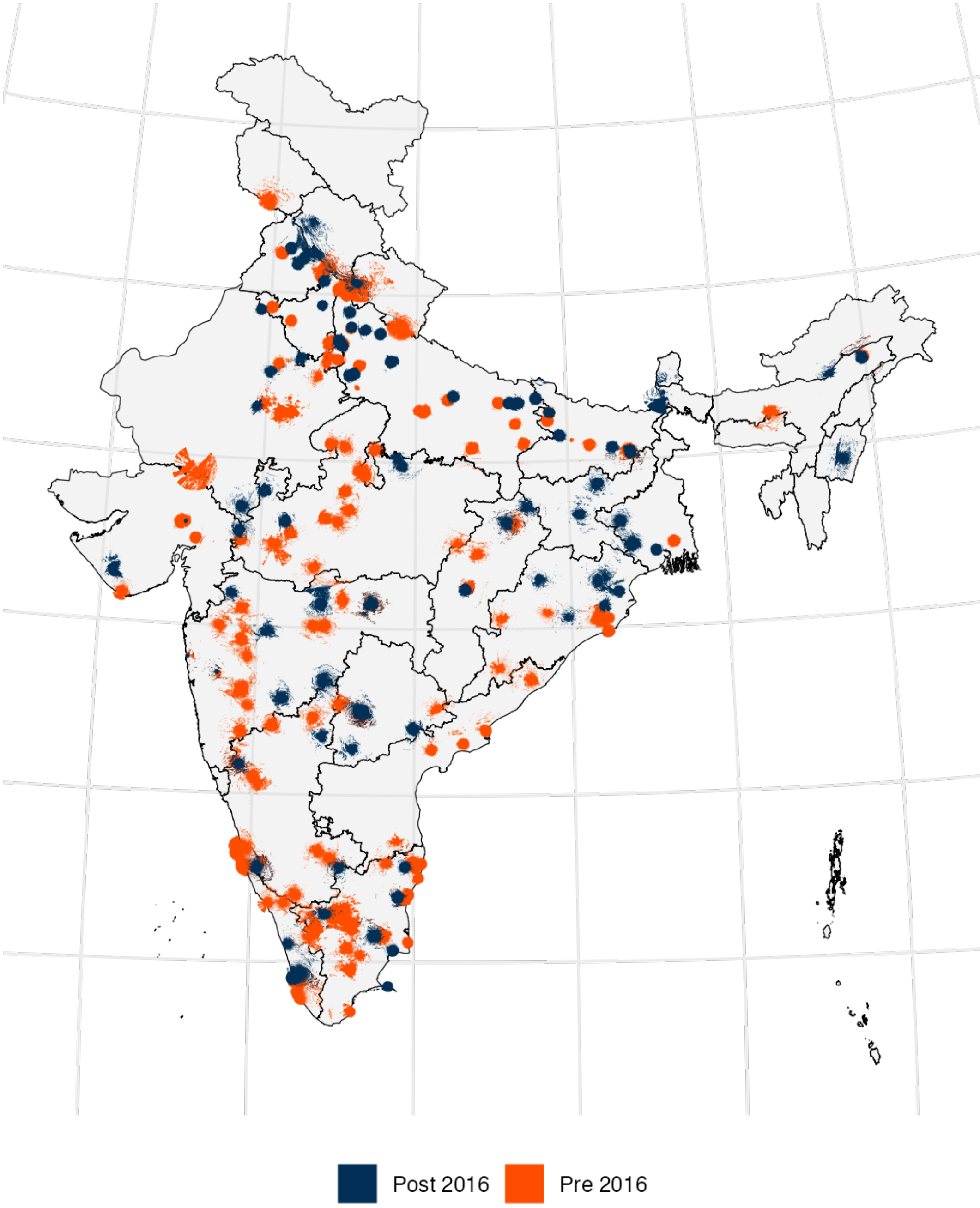


Figure B.5: Visualization of coverage areas of all 264 radio stations

Note: The above graph shows the coverage areas of all 264 geolocated radio stations launched by 2020. Colors indicate whether radios are launched before (red) or after (blue) 2016

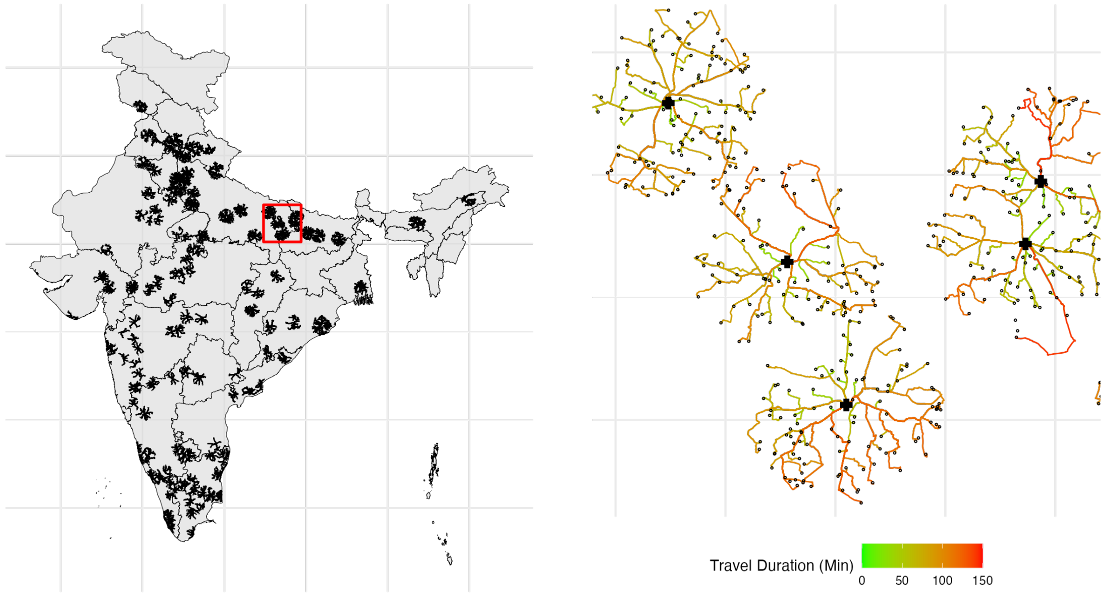


Figure B.6: Visualization of travel time from DHS observation to radio Tower

Note: the above map visualizes the data on travel times from the radio tower to the observation. The first map shows all travel routes obtained through Google Directions API. The map below shows a more detailed picture of the area in the red box above, showing travel routes, colored by travel time, from each DHS location in the vicinity of the radio station. The crosses indicate the radio tower locations. The dots indicate the locations of cluster observations as reported by the NFHS.

Variable	Description	Source
Radio Variables		
Exposure	Exposure to radio signal	own data and estimates
Coverage Probability	Probability of true location to lie in coverage area	own data and estimates
Radio Owner	Age 15 to 49: Household owns a radio	NFHS (women survey)
Radio Consumer	Age 15 to 54: Individual listens to radio	NFHS (women survey)
Radio Familyplanning	Age 15 to 54: Individual heard family planning message on radio in last few months	NFHS (women & men survey)
Radio HIV/AIDS	Age 15 to 49: Individual learned about AIDS from source: RADIO	NFHS (women survey)
Outcomes		
Years of Edu.	Years of education completed	NFHS (HH member survey)
Completed Primary	Completed primary school	NFHS (HH member survey)
Completed Secondary	Completed secondary school	NFHS (HH member survey)
Higher than Secondary	Education level higher than secondary school	NFHS (HH member survey)
Attends School	Age 5 to 18: Currently in School	NFHS (HH member survey)
Ever Married	Age >12: Was ever married (incl. divorced, widowed, married)	NFHS (HH member survey)
Num. Children	Age 15 to 49: Number of living children	NFHS (women survey)
Has Child	Age 15 to 49: Has at least one child that is alive	NFHS (women survey: state module)
Attitude (Count)	Age 15 to 49: Number of reasons that individual argues justify that a husband beats or hits his wife (0 to 5)	NFHS (women survey: state module)
Attitude (Any)	Age 15 to 49: Argues that husband is justified in hitting or beating his wife for at least on reason (out of 5)	NFHS (women survey: state module)
Autonomy	Married, Age 15 to 49: Share of decisions and places respondent participates in / can visit alone	NFHS (women survey)
Autonomy (Men)	Age 15 to 54: Share of decisions respondent believes his wife/partner should participate in	NFHS (men survey)
Any Violence (Ever)	Married, Age 15 to 49: Ever experienced any violence from partner (physical, emotional, sexual)	NFHS (women survey: state module)
Any Violence (past 12m)	Married, Age 15 to 49: Ever experienced any violence from partner (physical, emotional, sexual)	NFHS (women survey: state module)
Controls: Demography		
Age	Age of individual	NFHS (HH member survey)
Female	Individual is female	NFHS (HH member survey)
Caste ST	Individual is part of a Scheduled Tribe (inferred from caste of HH head)	NFHS (HH member survey)
Caste: SC	Individual is part of a Scheduled Caste (inferred from caste of HH head)	NFHS (HH member survey)
Caste: OBC	Individual is part of a Caste classified as Other Backward Caste (inferred from caste of HH head)	NFHS (HH member survey)
Caste: Other	Individual is part of another caste (inferred from caste of HH head)	NFHS (HH member survey)
Religion: Hindu	Individual is Hindu (inferred from religion of HH head)	NFHS (HH member survey)
Religion: Muslim	Individual is Muslim (inferred from religion of HH head)	NFHS (HH member survey)
Religion: Other	Individual is Other (inferred from religion of HH head)	NFHS (HH member survey)
Controls: Propagation		
Travel Time to Radio Tower (min)	Travel time (by car) to nearest radio tower that launched pre-2016	DHS locations & Google Directions API
Distance to Radio Tower (m)	Distance to nearest radio tower that launched before 2016	DHS locations & own data/ estimates
Expected Distance to Radio Tower (m)	Expected distance to nearest radio tower that launched before 2016	DHS locations & own data/ estimates
Mean Altitude	Mean altitude of 5km area surrounding observation	own estimates based on Jarvis et al. (2008)
Mean Ruggedness	Mean ruggedness of 5km area surrounding observation	own estimates based on Jarvis et al. (2008)
Additional Geographic Controls		
Urban	Cluster is classified as urban	
Pop. Density (2015)	Population density in	DHS Geospatial Covariate Dataset
Travel Time to Nearest City	Avg. time (minutes) required to reach the nearest high-density urban center	DHS Geospatial Covariate Dataset & Weiss et al. (2018)
Proximity: Water (m)	Geodesic distance to either a lake or the coastline	DHS Geospatial Covariate Dataset
Proximity: National Borders (m)	geodesic distance to the nearest international borders	DHS Geospatial Covariate Dataset

Table B.1: Variable descriptions and sources

C Functional Form: Explorative Analysis

In Equation 12, radio exposure is expected to exhibit a linear effect over time. To explore which alternative functional form may fit the regressions, I first define $Exposure_{i,m} = \sum_{r=1}^{R} AddedCoverageProbability_{i,R} \times Timeshare_R^m$ for $m \epsilon \{1, 2, 3\}$. Next, I run the regression in Equation 12 while adding all three exposure variables. I then take the derivate with respect to the timeshare and plot. The derivative differs for any variable y_i. To get an idea of the functional form, I focus on the effect of CRS on having listened to a family planning message on radio, an outcome that clearly relates to both listening to radio and the radios' topics.

Figure C.1 shows the resulting graph. Specifically, it plots the linear function as used in the paper and the polynomial described above. The polynomial appears to closely follow a quadratic form. Access to radio appears to have some immediate effects, which increasingly get stronger over time. These further closely resemble a quadratic functional form. Given that the polynomial is difficult to analyze in a table, for instance providing little information on whether effects are significant, I complement the linear effect by instead assuming that the effect of radio is quadratic over time, i.e. $Exposure_{i,2} = \sum_{r=1}^{R} AddedCoverageProbability_{i,R} \times Timeshare_R^2$. As the Figure shows, this graph closely follows that of the Polynomial. I, thus, report all results for both a linear and quadratic functional form in the paper.

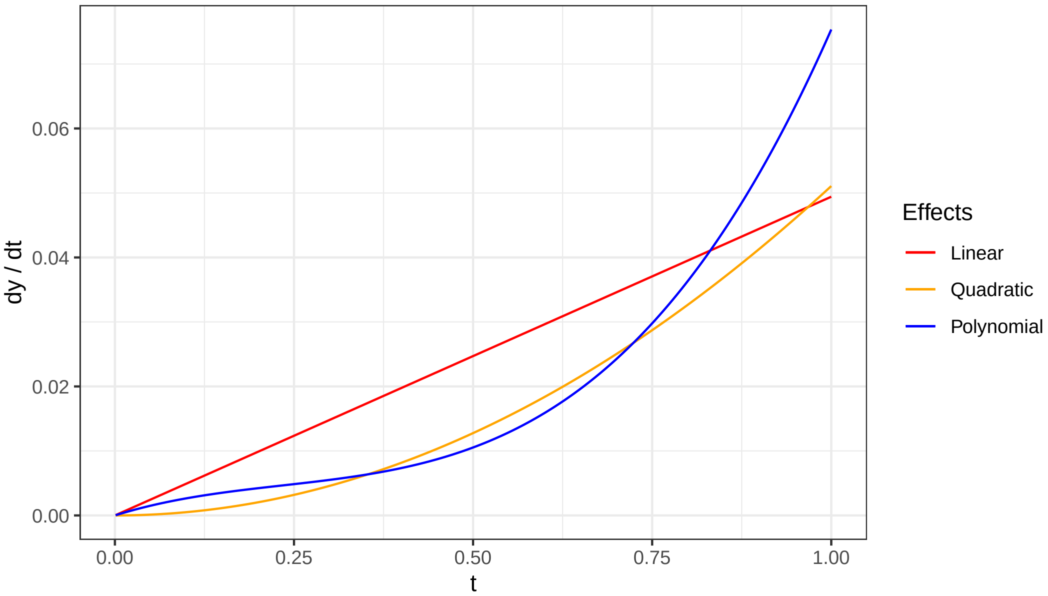


Figure C.1: Exploring non-linearity in treatment effects over time

D Additional Results

Panel A: Linear Effects Over Time

	Radio Consumer	Radio Familiyplanning	Radio: HIV/AIDS
Female x exposure	0.030	0.053**	0.077***
	(0.019)	(0.024)	(0.029)
Male x exposure	0.032	0.029	0.064**
	(0.035)	(0.033)	(0.028)
Num.Obs.	228 215	228 215	55 484
R2 Adj.	0.073	0.098	0.092
Mean Y	0.184	0.197	0.164
Mean Y: Female	0.165	0.191	0.149
Mean Y: Male	0.303	0.231	0.177
SD Y	0.388	0.397	0.37
Controls	✓	✓	✓
Radio FE	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Radio Consumer	Radio Familiyplanning	Radio: HIV/AIDS
Female x exposure2	0.047**	0.063**	0.085**
	(0.020)	(0.025)	(0.036)
Male x exposure2	0.057	0.040	0.072**
	(0.040)	(0.041)	(0.030)
Num.Obs.	228 215	228 215	55 484
R2 Adj.	0.073	0.098	0.092
Mean Y	0.184	0.197	0.164
Mean Y: Female	0.165	0.191	0.149
Mean Y: Male	0.303	0.231	0.177
SD Y	0.388	0.397	0.37
Controls	✓	✓	✓
Radio FE	✓	✓	✓

Table D.1: Exposure and radio consumption by gender

Note: The table shows the regression of radio consumption related variables on exposure. Regressions include all controls mentioned in Chapter V.1. The dependent variables are defined as follows: radio owner: household owns a radio; radio consumer: dummy indicating whether individual listens to radio at least less than once a week; radio family planning: dummy for whether individual heard a family planning message in last few months. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: Linear Effects Over Time

	Radio Consumer	Radio Intensity (0-3)	Radio: None (0)	Radio: (Almost) Daily (3)	Radio: At Least Weekly (2)	Radio: Less Than Weekly (1)
exposure	0.032*	0.055	−0.032*	0.007	0.009	0.016*
	(0.019)	(0.035)	(0.019)	(0.007)	(0.008)	(0.009)
Num.Obs.	196 472	196 472	196 472	196 472	196 472	196 472
R2 Adj.	0.060	0.063	0.060	0.041	0.019	0.019
Mean Y	0.165	0.322	0.835	0.047	0.063	0.055
SD Y	0.371	0.792	0.371	0.212	0.242	0.229
Controls	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Radio Consumer	Radio Intensity (0-3)	Radio: None (0)	Radio: (Almost) Daily (3)	Radio: At Least Weekly (2)	Radio: Less Than Weekly (1)
exposure2	0.052***	0.102**	−0.052***	0.015	0.019**	0.018*
	(0.020)	(0.045)	(0.020)	(0.013)	(0.009)	(0.009)
Num.Obs.	196 472	196 472	196 472	196 472	196 472	196 472
R2 Adj.	0.060	0.063	0.060	0.041	0.019	0.019
Mean Y	0.165	0.322	0.835	0.047	0.063	0.055
SD Y	0.371	0.792	0.371	0.212	0.242	0.229
Controls	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓

Table D.2: Intensity of radio consumption

Note: The table shows the regression of radio consumption related variables on exposure. Regressions include all controls mentioned in Chapter V.1. The dependent variables are defined as follows: radio consumer: individual listens to radio; radio intensity: ordinal scale of intensity ranging from not at all (0) to (almost) daily (3). The following columns are four indicator variables for each level of intensity. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

(A) Child Does Not Go to School

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure	−0.015**	−0.002	−0.001	−0.035*	−0.059***
	(0.007)	(0.002)	(0.008)	(0.019)	(0.022)
Male x exposure	−0.007	−0.001	−0.004	−0.024*	−0.025*
	(0.005)	(0.003)	(0.005)	(0.014)	(0.015)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.087	0.006	0.046	0.162	0.312

(B) Reason: Interest

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure	−0.008**	0.000	−0.001	−0.026***	−0.020
	(0.004)	(0.001)	(0.005)	(0.009)	(0.015)
Male x exposure	−0.004	−0.001	−0.005	−0.019*	−0.001
	(0.003)	(0.001)	(0.005)	(0.010)	(0.014)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.028	0.002	0.017	0.056	0.094

(C) Reason: Costs too High

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure	−0.005	−0.001	−0.002	−0.003	−0.026*
	(0.004)	(0.001)	(0.003)	(0.013)	(0.015)
Male x exposure	−0.006	0.000	−0.004*	−0.010	−0.022
	(0.004)	(0.001)	(0.003)	(0.012)	(0.015)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.017	0.001	0.01	0.032	0.058

(D) Reason: Marriage

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure	−0.002**	0.000	0.000	0.000	−0.020**
	(0.001)	(0.000)	(0.000)	(0.003)	(0.010)
Male x exposure	0.001	0.000	0.000	0.001	0.003
	(0.001)	(0.000)	(0.000)	(0.002)	(0.008)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.004	0	0	0.003	0.026

(E) Reason: Mostly Female-specific Household and care work, no school for girls available, not safe, no female teacher

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure	0.001	−0.001	0.003	−0.002	0.004
	(0.003)	(0.001)	(0.003)	(0.009)	(0.010)
Male x exposure	0.001	−0.001	0.003*	0.008	−0.008
	(0.003)	(0.001)	(0.002)	(0.007)	(0.011)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.014	0.001	0.008	0.027	0.049

(F) Reason: Work Work in Family Business or Outside Home

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure	0.001	0.000	0.000	0.006	−0.001
	(0.002)	(0.000)	(0.002)	(0.006)	(0.007)
Male x exposure	−0.001	0.000	0.001	−0.008	0.001
	(0.002)	(0.000)	(0.002)	(0.005)	(0.009)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.005	0	0.003	0.009	0.021

(G) Reason: Availability Too far away, Transport

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure	0.000	−0.001	0.001	−0.012*	0.010
	(0.002)	(0.001)	(0.002)	(0.006)	(0.008)
Male x exposure	0.002	−0.001	0.003	−0.002	0.009
	(0.002)	(0.001)	(0.002)	(0.005)	(0.007)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.006	0.001	0.003	0.013	0.02

(H) Reason: Other Not Necessary, Failure, Not Admitted to School, Other

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure	0.000	0.001	−0.001	0.002	−0.007
	(0.003)	(0.001)	(0.002)	(0.007)	(0.010)
Male x exposure	0.000	0.001	−0.001	0.005	−0.008
	(0.003)	(0.001)	(0.002)	(0.010)	(0.010)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.012	0.002	0.005	0.022	0.045

Table D.3: Linear model – Reasons for not going to school

Note: The table shows regressions of reasons for not going to school on exposure interacted with a child's gender. Table (A) is an indicator for not going to school. All other variables are indicators for whether a child dropped out of school for the specified reason. The variable is defined as zero for all children still going to school at the time of the survey and for those dropping out for a different reason. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

(A) Child Does Not Go to School

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure2	−0.021***	−0.005**	−0.007	−0.036	−0.080***
	(0.008)	(0.003)	(0.007)	(0.024)	(0.024)
Male x exposure2	−0.007	−0.002	−0.004	−0.017	−0.036**
	(0.005)	(0.003)	(0.005)	(0.014)	(0.016)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.087	0.006	0.046	0.162	0.312

(B) Reason: Interest

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure2	−0.006*	0.000	−0.001	−0.019	−0.018
	(0.004)	(0.001)	(0.004)	(0.013)	(0.015)
Male x exposure2	−0.005*	0.000	−0.006	−0.023**	−0.007
	(0.003)	(0.001)	(0.004)	(0.010)	(0.015)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.028	0.002	0.017	0.056	0.094

(C) Reason: Costs too High

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure2	−0.007**	−0.002	−0.004	−0.008	−0.027**
	(0.004)	(0.001)	(0.003)	(0.012)	(0.012)
Male x exposure2	−0.006**	0.000	−0.005	−0.011*	−0.018
	(0.003)	(0.001)	(0.003)	(0.006)	(0.015)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.017	0.001	0.01	0.032	0.058

(D) Reason: Marriage

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure2	−0.004**	0.000	0.000	−0.003	−0.026***
	(0.001)	(0.000)	(0.000)	(0.003)	(0.010)
Male x exposure2	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.000)	(0.000)	(0.002)	(0.008)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.004	0	0	0.003	0.026

(E) Reason: Mostly Female-specific Household and care work, no school for girls available, not safe, no female teacher

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure2	0.000	−0.001	0.001	−0.004	0.001
	(0.003)	(0.000)	(0.003)	(0.008)	(0.012)
Male x exposure2	0.001	−0.001	0.003	0.013*	−0.013
	(0.002)	(0.001)	(0.002)	(0.007)	(0.010)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.014	0.001	0.008	0.027	0.049

(F) Reason: Work Work in Family Business or Outside Home

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure2	0.001	0.000	0.001	0.009	−0.001
	(0.002)	(0.000)	(0.002)	(0.007)	(0.007)
Male x exposure2	0.001	0.000	0.001	−0.009**	0.009
	(0.002)	(0.000)	(0.002)	(0.005)	(0.010)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.005	0	0.003	0.009	0.021

(G) Reason: Availability Too far away, Transport, Not Admitted to School

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure2	0.000	−0.001**	0.000	−0.009	0.011
	(0.002)	(0.000)	(0.002)	(0.006)	(0.011)
Male x exposure2	0.002	−0.001**	0.003	0.002	0.008
	(0.002)	(0.001)	(0.002)	(0.005)	(0.007)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.006	0.001	0.003	0.013	0.02

(H) Reason: Other Not Necessary, Failure, Other

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure2	−0.004	−0.001	−0.003	−0.002	−0.020*
	(0.003)	(0.001)	(0.002)	(0.009)	(0.011)
Male x exposure2	0.000	0.001	−0.001	0.011	−0.016
	(0.003)	(0.002)	(0.002)	(0.009)	(0.010)
Num.Obs.	204 342	83 093	60 519	30 262	30 468
Mean Y	0.012	0.002	0.005	0.022	0.045

Table D.4: Quadratic model – reasons for not going to school (Quadratic Effects over Time)

Note: The table shows regressions of reasons for not going to school on exposure with quadratic effects over time interacted with a child's gender. Table (A) is an indicator for not going to school. All other variables are indicators for whether a child dropped out of school for the specified reason. The variable is defined as zero for all children still going to school at the time of the survey and for those dropping out for a different reason. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: Linear Effects Over Time

	Married (13-18)	Married (19-24)	Married (25-29)	Married (30-34)	Married (35-39)
Female x exposure	−0.014**	−0.049**	−0.027	0.008	0.006
	(0.007)	(0.019)	(0.024)	(0.011)	(0.008)
Male x exposure	0.009	−0.012	−0.043**	−0.018	−0.002
	(0.006)	(0.018)	(0.020)	(0.012)	(0.009)
Num.Obs.	95 333	87 440	68 221	57 062	46 466
R2 Adj.	0.060	0.284	0.193	0.069	0.019
Mean Y	0.038	0.388	0.76	0.924	0.968
Mean Y: Female	0.06	0.573	0.898	0.971	0.984
Mean Y: Male	0.019	0.195	0.616	0.876	0.953
SD Y	0.192	0.487	0.427	0.265	0.175
Controls	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Married (13-18)	Married (19-24)	Married (25-29)	Married (30-34)	Married (35-39)
Female x exposure2	−0.022**	−0.062**	−0.042	0.010	0.000
	(0.009)	(0.024)	(0.029)	(0.010)	(0.009)
Male x exposure2	0.010	−0.019	−0.062**	−0.016	−0.007
	(0.007)	(0.022)	(0.024)	(0.011)	(0.010)
Num.Obs.	95 333	87 440	68 221	57 062	46 466
R2 Adj.	0.060	0.284	0.193	0.069	0.019
Mean Y	0.038	0.388	0.76	0.924	0.968
Mean Y: Female	0.06	0.573	0.898	0.971	0.984
Mean Y: Male	0.019	0.195	0.616	0.876	0.953
SD Y	0.192	0.487	0.427	0.265	0.175
Controls	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓

Table D.5: Effect of community radio stations on marriage status

Note: The tables show separate regressions for whether the person surveyed has ever been married by age cohort on exposure to radio. Panel A assumes linear effects over time and Panel B quadratic effects. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: Linear Effects Over Time

	Autonomy Decision	Autonomy Decision (15-25)	Autonomy Decision (26-35)	Autonomy Decision (36-45)	Autonomy Decision (45-49)
exposure	0.042	0.120***	0.036	−0.013	0.078
	(0.030)	(0.040)	(0.041)	(0.030)	(0.068)
Num.Obs.	24 400	5481	9571	7210	2138
R2 Adj.	0.058	0.065	0.035	0.040	0.036
Mean Y	0.741	0.655	0.744	0.786	0.793
SD Y	0.382	0.419	0.378	0.355	0.355
Controls	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Autonomy Decision	Autonomy Decision (15-25)	Autonomy Decision (26-35)	Autonomy Decision (36-45)	Autonomy Decision (45-49)
exposure2	0.030	0.129***	0.024	−0.026	0.073
	(0.037)	(0.047)	(0.047)	(0.033)	(0.077)
Num.Obs.	24 400	5481	9571	7210	2138
R2 Adj.	0.058	0.064	0.035	0.040	0.036
Mean Y	0.741	0.655	0.744	0.786	0.793
SD Y	0.382	0.419	0.378	0.355	0.355
Controls	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓

Table D.6: Autonomy of women (decisions)

Note: The tables show separate regressions of the share of decisions a woman participates in on exposure to radio. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: Linear Effects Over Time

	Autonomy Mobility	Autonomy Mobility (15-25)	Autonomy Mobility (26-35)	Autonomy Mobility (36-45)	Autonomy Mobility (45-49)
exposure	0.034	0.082**	−0.013	0.012	0.062
	(0.028)	(0.040)	(0.029)	(0.048)	(0.065)
Num.Obs.	34 302	13 435	10 385	8000	2482
R2 Adj.	0.165	0.114	0.116	0.097	0.123
Mean Y	0.506	0.368	0.545	0.637	0.667
SD Y	0.456	0.436	0.454	0.435	0.429
Controls	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Autonomy Mobility	Autonomy Mobility (15-25)	Autonomy Mobility (26-35)	Autonomy Mobility (36-45)	Autonomy Mobility (45-49)
exposure2	0.016	0.082*	−0.028	−0.032	0.033
	(0.028)	(0.046)	(0.033)	(0.043)	(0.076)
Num.Obs.	34 302	13 435	10 385	8000	2482
R2 Adj.	0.165	0.114	0.117	0.097	0.123
Mean Y	0.506	0.368	0.545	0.637	0.667
SD Y	0.456	0.436	0.454	0.435	0.429
Controls	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓

Table D.7: Autonomy of women (mobility)

Note: The table shows separate regressions of the share of places a woman can visit on her own by age cohort on exposure to radio. Panel A assumes linear effects over time and Panel B quadratic effects. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: Linear Effects Over Time

	Attitude Any	Attitude Any	Attitude Count	Attitude Count
exposure	−0.043		−0.102	
	(0.038)		(0.129)	
15-24 x exposure		−0.059		−0.121
		(0.043)		(0.151)
25-34 x exposure		−0.040		−0.128
		(0.036)		(0.126)
35-44 x exposure		−0.040		−0.095
		(0.038)		(0.118)
45-49 x exposure		−0.005		0.024
		(0.050)		(0.165)
Num.Obs.	33 420	33 420	33 420	33 420
R2 Adj.	0.146	0.146	0.129	0.129
Mean Y	0.406	0.406	1.098	1.098
SD Y	0.491	0.491	1.606	1.606
Controls	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Attitude Any	Attitude Any	Attitude Count	Attitude Count
exposure2	−0.042		−0.106	
	(0.043)		(0.134)	
15-24 x exposure2		−0.058		−0.103
		(0.049)		(0.163)
25-34 x exposure2		−0.042		−0.134
		(0.041)		(0.131)
35-44 x exposure2		−0.036		−0.127
		(0.041)		(0.112)
45-49 x exposure2		−0.011		0.019
		(0.067)		(0.203)
Num.Obs.	33 420	33 420	33 420	33 420
R2 Adj.	0.146	0.146	0.129	0.129
Mean Y	0.406	0.406	1.098	1.098
SD Y	0.491	0.491	1.606	1.606
Controls	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓

Table D.8: Attitudes of women towards domestic violence

Note: The above table regresses a variable for whether women in the NFHS's domestic violence sample agree that men are justified to beat their wife under a surveyed circumstances. These include: going out without telling husband, neglecting children, arguing with husband, refusing to have sex, or improper cooking. *Attitude Any* is a dummy for whether the woman agrees with any of the reasons. *Attitude Count* is an additive variable for the number of reasons a woman agrees with. Data on domestic violence stems from the NFHS's state module, which is carried out in 15% households and 30% of clusters and substantially longer than the standard questionnaire. In each selected household, a random woman above the age of 15 was selected for the survey. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: Linear Effects Over Time

	Attitude Any	Attitude Any	Attitude Count	Attitude Count
exposure	−0.043		−0.102	
	(0.038)		(0.129)	
15-24 x exposure		−0.059		−0.121
		(0.043)		(0.151)
25-34 x exposure		−0.040		−0.128
		(0.036)		(0.126)
35-44 x exposure		−0.040		−0.095
		(0.038)		(0.118)
45-49 x exposure		−0.005		0.024
		(0.050)		(0.165)
Num.Obs.	33 420	33 420	33 420	33 420
R2 Adj.	0.146	0.146	0.129	0.129
Mean Y	0.406	0.406	1.098	1.098
SD Y	0.491	0.491	1.606	1.606
Controls	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Attitude Any	Attitude Any	Attitude Count	Attitude Count
exposure2	0.023		0.075	
	(0.045)		(0.126)	
15-24 x exposure2		0.012		0.054
		(0.051)		(0.155)
25-34 x exposure2		0.034		0.037
		(0.042)		(0.110)
35-44 x exposure2		0.022		0.123
		(0.044)		(0.124)
45-49 x exposure2		−0.031		−0.186
		(0.046)		(0.123)
Num.Obs.	30 961	30 961	30 961	30 961
R2 Adj.	0.102	0.102	0.088	0.088
Mean Y	0.303	0.303	0.696	0.696
SD Y	0.46	0.46	1.281	1.281
Controls	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓

Table D.9: Men's attitudes toward domestic violence

Note: The above table regresses a variable for whether men agree that a husband is justified to beat his wife under a surveyed circumstances. These include: going out without telling husband, neglecting children, arguing with husband, refusing to have sex, or improper cooking. Attitude (Any) is a dummy for whether the man agrees with any of the reasons. Attitude (Count) describes the number of reasons the respondent agreed with. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: Linear Effects Over Time

	Any Violence Ever	Any Violence Ever	Any Violence 12m	Any Violence 12m	Sexual Viol. 12m	Sexual Viol. 12m	Emotional Viol. 12m	Emotional Viol. 12m	Physical Viol. 12m	Physical Viol. 12m
exposure	−0.027		−0.007		0.014		0.017		−0.026	
	(0.034)		(0.031)		(0.020)		(0.027)		(0.028)	
15-24 x exposure		−0.044		−0.035		−0.012		0.031		−0.052
		(0.052)		(0.047)		(0.028)		(0.044)		(0.042)
25-34 x exposure		−0.033		−0.016		0.019		0.006		−0.038
		(0.038)		(0.037)		(0.023)		(0.031)		(0.031)
35-44 x exposure		−0.022		0.007		0.016		0.017		−0.016
		(0.036)		(0.032)		(0.019)		(0.029)		(0.032)
45-49 x exposure		0.004		0.026		0.023		0.036		0.020
		(0.043)		(0.037)		(0.023)		(0.033)		(0.034)
Num.Obs.	18 388	18 388	18 388	18 388	18 388	18 388	18 388	18 388	18 388	18 388
R2 Adj.	0.082	0.082	0.071	0.071	0.024	0.024	0.039	0.039	0.069	0.069
Mean Y	0.332	0.332	0.27	0.27	0.057	0.057	0.113	0.113	0.232	0.232
SD Y	0.471	0.471	0.444	0.444	0.232	0.232	0.317	0.317	0.422	0.422
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Any Violence Ever	Any Violence Ever	Any Violence 12m	Any Violence 12m	Sexual Viol. 12m	Sexual Viol. 12m	Emotional Viol. 12m	Emotional Viol. 12m	Physical Viol. 12m	Physical Viol. 12m
exposure2	−0.051		−0.010		0.012		0.013		−0.037	
	(0.040)		(0.036)		(0.024)		(0.030)		(0.029)	
15-24 x exposure2		−0.085		−0.054		−0.029		0.017		−0.083*
		(0.061)		(0.060)		(0.031)		(0.059)		(0.048)
25-34 x exposure2		−0.070		−0.034		0.021		0.004		−0.069**
		(0.044)		(0.042)		(0.027)		(0.034)		(0.032)
35-44 x exposure2		−0.019		0.028		0.013		0.019		0.002
		(0.044)		(0.040)		(0.021)		(0.038)		(0.035)
45-49 x exposure2		−0.025		0.028		0.027		0.020		0.024
		(0.052)		(0.048)		(0.033)		(0.038)		(0.045)
Num.Obs.	18 388	18 388	18 388	18 388	18 388	18 388	18 388	18 388	18 388	18 388
R2 Adj.	0.082	0.082	0.071	0.071	0.024	0.024	0.039	0.039	0.069	0.069
Mean Y	0.332	0.332	0.27	0.27	0.057	0.057	0.113	0.113	0.232	0.232
SD Y	0.471	0.471	0.444	0.444	0.232	0.232	0.317	0.317	0.422	0.422
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table D.10: Experience of domestic violence in past 12 months

Note: The above table regresses a variable for whether a woman in the NFHS's domestic violence sample experienced form of violence from her partner ever or in the past 12 months (Columns 1-4). Columns (5) to (10) show the different types of violence surveyed, including sexual, physical, and emotional violence. The outcome variables are binary. Data on domestic violence stems from the NFHS's state module, which is carried out in 15% households and 30% of clusters and substantially longer than the standard questionnaire. In each selected household, a random woman above the age of 15 was selected for the survey. Panel A assumes linear effects over time and Panel B quadratic effects. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: Linear Effects Over Time

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure	−0.005	0.000	0.039*	0.058***
	(0.012)	(0.011)	(0.022)	(0.020)
Male x exposure	0.000	0.003	0.019	0.032*
	(0.014)	(0.010)	(0.015)	(0.016)
Num.Obs.	91 355	62 612	31 731	32 330
R2 Adj.	0.136	0.094	0.127	0.137
Mean Y	0.904	0.922	0.799	0.648
Mean Y: Female	0.9	0.916	0.779	0.612
Mean Y: Male	0.907	0.928	0.818	0.681
SD Y	0.295	0.267	0.4	0.478
Controls	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
Female x exposure2	−0.002	0.005	0.045	0.088***
	(0.012)	(0.011)	(0.027)	(0.022)
Male x exposure2	−0.002	0.007	0.015	0.054***
	(0.016)	(0.009)	(0.016)	(0.018)
Num.Obs.	91 355	62 612	31 731	32 330
R2 Adj.	0.136	0.094	0.127	0.138
Mean Y	0.904	0.922	0.799	0.648
Mean Y: Female	0.9	0.916	0.779	0.612
Mean Y: Male	0.907	0.928	0.818	0.681
SD Y	0.295	0.267	0.4	0.478
Controls	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓

Table D.11: Exposure and school attendance by age group

Note: The dependent variable in the above regressions indicates whether an individual in a given age group attended school at the time of the survey. The variable is only collected for children up to the age of 18. Panel A assumes linear effects over time and Panel B quadratic effects. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: Linear Effects Over Time

	TV Owner	TV Consumer	Newspaper Consumer	Internet User	Mobile Phone Owner
exposure	−0.017	−0.008	−0.005	0.013	0.030
	(0.022)	(0.019)	(0.018)	(0.015)	(0.034)
Num.Obs.	190 090	196 472	196 472	167 086	34 302
R2 Adj.	0.295	0.237	0.189	0.174	0.213
Mean Y	0.752	0.817	0.45	0.148	0.494
SD Y	0.432	0.387	0.498	0.355	0.5
Controls	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	TV Owner	TV Consumer	Newspaper Consumer	Internet User	Mobile Phone Owner
exposure2	0.016	0.016	0.005	0.019	0.028
	(0.029)	(0.026)	(0.023)	(0.019)	(0.033)
Num.Obs.	190 090	196 472	196 472	167 086	34 302
R2 Adj.	0.295	0.237	0.189	0.174	0.213
Mean Y	0.752	0.817	0.45	0.148	0.494
SD Y	0.432	0.387	0.498	0.355	0.5
Controls	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓

Table D.12: Exposure and non-radio media

Note: The above regressions test whether treatment affects other types of media consumption. This includes whether (1) household has a TV, (2) a dummy indicating whether individual watches TV or (3) reads the newspaper at least less than once week, (4) the household has access to internet, or (5) owns a mobile phone. Panel A assumes linear effects over time and Panel B quadratic effects. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: Linear Effects Over Time

	Caste SC/ST	Muslim	Urban	Log. Pop. Density	Log. Travel Time City (min)	Proximity Borders (m)	Travel Time Radio (min)
exposure	−0.014	0.026	0.054	0.311	0.033	817.310	3.585
	(0.022)	(0.019)	(0.079)	(0.249)	(0.191)	(2873.507)	(2.688)
Num.Obs.	167 086	171 878	171 878	171 878	171 878	171 878	171 878
R2 Adj.	0.075	0.100	0.363	0.793	0.586	0.985	0.825
Distance Controls	✓	✓	✓	✓	✓	✓	✓
Geography Controls	✓	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓	✓

Panel B: Quadratic Effects Over Time

	Caste SC/ST	Muslim	Urban	Log. Pop. Density	Log. Travel Time City (min)	Proximity Borders (m)	Travel Time Radio (min)
exposure2	−0.037	0.005	0.022	0.306	0.059	421.987	2.903
	(0.025)	(0.022)	(0.088)	(0.266)	(0.225)	(2864.688)	(2.783)
Num.Obs.	167 086	171 878	171 878	171 878	171 878	171 878	171 878
R2 Adj.	0.075	0.100	0.363	0.792	0.586	0.985	0.825
Distance Controls	✓	✓	✓	✓	✓	✓	✓
Geography Controls	✓	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓	✓

Table D.13: Exogeneity check: correlation of treatment variation with observables

Note: The table show regressions of different covariates that are unlikely to be affected by radio on radio exposure. Regressions control for propagation controls and CRS dummies only. Regressions on travel time to the nearest radio station additionally exclude this variable from the set of propagation controls. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

E Results without Jittering Correction

Panel A.1: Linear Effects Over Time - Reported Distance

	Radio Owner	Radio Consumer	Radio Familyplanning	Radio HIV/AIDS
exposure (point)	−0.002	0.012	0.030**	0.047**
	(0.011)	(0.012)	(0.014)	(0.020)
Num.Obs.	190 157	228 289	228 289	55 508
R2 Adj.	0.065	0.073	0.098	0.092
Mean Y	0.095	0.184	0.197	0.164

Panel B.1: Quadratic Effects Over Time - Reported Distance

	Radio Owner	Radio Consumer	Radio Familyplanning	Radio HIV/AIDS
exposure2 (point)	−0.001	0.035	0.050**	0.076***
	(0.012)	(0.021)	(0.022)	(0.027)
Num.Obs.	190 157	228 289	228 289	55 508
R2 Adj.	0.065	0.073	0.098	0.093
Mean Y	0.095	0.184	0.197	0.164

Panel A.2: Linear Effects Over Time - Expected Distance

	Radio Owner	Radio Consumer	Radio Familyplanning	Radio HIV/AIDS
exposure (point)	−0.003	0.014	0.032**	0.047**
	(0.011)	(0.013)	(0.014)	(0.020)
Num.Obs.	190 157	228 289	228 289	55 508
R2 Adj.	0.065	0.073	0.098	0.092
Mean Y	0.095	0.184	0.197	0.164

Panel B.2: Quadratic Effects Over Time - Expected Distance

	Radio Owner	Radio Consumer	Radio Familyplanning	Radio HIV/AIDS
exposure2 (point)	−0.002	0.039*	0.053**	0.079***
	(0.012)	(0.022)	(0.023)	(0.027)
Num.Obs.	190 157	228 289	228 289	55 508
R2 Adj.	0.065	0.073	0.098	0.093
Mean Y	0.095	0.184	0.197	0.164

Table E.1: Radio consumption: evaluation of jittering algorithm

Note: Panels report the results using either a linear or quadratic effect over time. This is done using the reported location ("exposure (point)"). Panel A.1 and A.2 use distance controls relying on the "Reported Distance", i.e., solely relying on the distance between the jittered location and the radio tower. Panels A.2 and B.2 correct this distance by taking the jittering into account, controlling for expected distances. The dependent variables are defined as follows: radio owner: household owns a radio; radio consumer: dummy indicating whether individual listens to radio at least less than once a week; radio family planning: dummy for whether individual heard a family planning message in last few months. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: years of education

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-19)	Higher Education (19-30)	All (5-30)
Female x Exposure	0.070	0.245***	0.392**	0.282	0.493**	0.309**
	(0.064)	(0.093)	(0.162)	(0.189)	(0.219)	(0.122)
Male x Exposure	0.051	0.223***	0.139	0.121	0.195	0.178
	(0.059)	(0.082)	(0.117)	(0.189)	(0.199)	(0.120)
Num.Obs.	91 341	62 587	31 705	45 395	174 402	392 353
R2 Adj.	0.637	0.345	0.195	0.186	0.233	0.534
Mean Y	1.68	5.941	8.345	9.66	9.458	6.996
SD Y	1.635	2.116	2.584	3.406	4.912	4.855

Panel B: degree obtained

	Primary	Secondary	Higher
Female x Exposure	0.032**	0.037*	0.027*
	(0.015)	(0.021)	(0.015)
Male x Exposure	0.013	0.017	0.015
	(0.013)	(0.021)	(0.016)
Num.Obs.	238 425	191 899	191 899
R2 Adj.	0.161	0.160	0.134
Mean Y	0.807	0.41	0.255
SD Y	0.395	0.492	0.436

Panel C: is married

	Married (13-18)	Married (19-24)	Married (25-29)	Married (30-34)	Married (35-39)
Female x Exposure	−0.015**	−0.051***	−0.025	0.009	0.005
	(0.007)	(0.020)	(0.026)	(0.012)	(0.008)
Male x Exposure	0.009	−0.015	−0.042**	−0.017	−0.004
	(0.006)	(0.017)	(0.021)	(0.012)	(0.009)
Num.Obs.	95 359	87 467	68 256	57 081	46 469
R2 Adj.	0.060	0.284	0.193	0.069	0.019
Mean Y	0.038	0.388	0.76	0.924	0.968
SD Y	0.192	0.487	0.427	0.265	0.175

Panel D: number of children

	# Children (15-18)	# Children (19-25)	# Children (26-30)	# Children (31-35)	# Children (36-40)	# Children (41-49)
exposure	−0.001	−0.079**	−0.138**	−0.210**	−0.033	−0.012
	(0.003)	(0.034)	(0.069)	(0.098)	(0.074)	(0.104)
Num.Obs.	20 747	56 848	32 510	26 469	24 899	35 064
R2 Adj.	0.011	0.306	0.198	0.232	0.254	0.282
Mean Y	0.006	0.624	1.882	2.429	2.735	2.993
SD Y	0.084	0.899	1.199	1.295	1.43	1.633

Panel E: autonomy of women

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-49)
exposure	0.037*	0.120***	0.009	−0.001	0.068
	(0.021)	(0.034)	(0.024)	(0.028)	(0.064)
Num.Obs.	24 411	5484	9572	7212	2143
R2 Adj.	0.147	0.138	0.097	0.091	0.106
Mean Y	0.635	0.505	0.639	0.704	0.718
SD Y	0.329	0.336	0.322	0.309	0.308

Table E.2: Main results with linear effects over time and assuming uniform distribution of survey locations $f(x^*)$

Note: The tables repeat the paper's main regressions with point exposure as the treatment variable. Further, expected distance controls are replaced by point distance controls. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: years of education

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-19)	Higher Education (19-30)	All (5-30)
is female = 1 x exposure2	0.062	0.252**	0.494***	0.522**	0.770***	0.461***
	(0.067)	(0.108)	(0.168)	(0.230)	(0.236)	(0.138)
is female = 0 x exposure2	0.053	0.259***	0.137	0.418**	0.552***	0.370***
	(0.057)	(0.076)	(0.107)	(0.201)	(0.195)	(0.122)
Num.Obs.	91 341	62 587	31 705	45 395	174 402	392 353
R2 Adj.	0.637	0.345	0.195	0.186	0.233	0.534
Mean Y	1.68	5.941	8.345	9.66	9.458	6.996

Panel B: degree obtained

	Primary	Secondary	Higher
is female = 1 x exposure2	0.050***	0.063***	0.044***
	(0.017)	(0.023)	(0.015)
is female = 0 x exposure2	0.031**	0.059***	0.052***
	(0.013)	(0.021)	(0.015)
Num.Obs.	238 425	191 899	191 899
R2 Adj.	0.161	0.160	0.134
Mean Y	0.807	0.41	0.255

Panel C: is married

	Married (13-18)	Married (19-24)	Married (25-29)	Married (30-34)	Married (35-39)
is female = 1 x exposure2	−0.023**	−0.061**	−0.043	0.010	−0.002
	(0.009)	(0.025)	(0.030)	(0.010)	(0.009)
is female = 0 x exposure2	0.009	−0.019	−0.064***	−0.018	−0.008
	(0.006)	(0.022)	(0.024)	(0.012)	(0.010)
Num.Obs.	95 359	87 467	68 256	57 081	46 469
R2 Adj.	0.060	0.284	0.193	0.068	0.019
Mean Y	0.038	0.388	0.76	0.924	0.968

Panel D: number of children

	# Children (15-18)	# Children (19-25)	# Children (26-30)	# Children (31-35)	# Children (36-40)	# Children (41-49)
exposure2	−0.001	−0.097**	−0.188***	−0.306***	−0.164	−0.141
	(0.003)	(0.038)	(0.064)	(0.085)	(0.111)	(0.097)
Num.Obs.	20 747	56 848	32 510	26 469	24 899	35 064
R2 Adj.	0.011	0.306	0.199	0.233	0.255	0.282
Mean Y	0.006	0.624	1.882	2.429	2.735	2.993

Panel E: autonomy of women

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-49)
exposure2	0.018	0.104**	−0.003	−0.028	0.065
	(0.027)	(0.050)	(0.031)	(0.027)	(0.068)
Num.Obs.	24 411	5484	9572	7212	2143
R2 Adj.	0.147	0.137	0.097	0.091	0.106
Mean Y	0.635	0.505	0.639	0.704	0.718

Table E.3: Main results with quadratic effects over time and assuming uniform distribution of survey locations $f(x^*)$

Note: The tables repeat the paper's main regressions with point exposure as the treatment variable. Further, expected distance controls are replaced by point distance controls. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: years of education

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-19)	Higher Education (19-30)	All (5-30)
is female = 1 x exposure (point)	0.062	0.143	0.219	0.158	0.290*	0.183*
	(0.038)	(0.095)	(0.135)	(0.139)	(0.165)	(0.105)
is female = 0 x exposure (point)	0.031	0.110	0.028	0.080	0.044	0.071
	(0.035)	(0.077)	(0.096)	(0.134)	(0.169)	(0.104)
Num.Obs.	91 341	62 587	31 705	45 395	174 402	392 353
R2 Adj.	0.637	0.345	0.194	0.185	0.232	0.534
Mean Y	1.68	5.941	8.345	9.66	9.458	6.996

Panel B: degree obtained

	Primary	Secondary	Higher
is female = 1 x exposure (point)	0.020*	0.021	0.019
	(0.011)	(0.018)	(0.013)
is female = 0 x exposure (point)	0.004	0.007	0.009
	(0.010)	(0.019)	(0.013)
Num.Obs.	238 425	191 899	191 899
R2 Adj.	0.161	0.160	0.134
Mean Y	0.807	0.41	0.255

Panel C: is married

	Married (13-18)	Married (19-24)	Married (25-29)	Married (30-34)	Married (35-39)
is female = 1 x exposure (point)	−0.006	−0.027*	−0.013	0.002	0.005
	(0.007)	(0.015)	(0.021)	(0.007)	(0.005)
is female = 0 x exposure (point)	0.014***	0.004	−0.028	−0.020**	0.001
	(0.005)	(0.012)	(0.017)	(0.009)	(0.008)
Num.Obs.	95 359	87 467	68 256	57 081	46 469
R2 Adj.	0.060	0.284	0.193	0.069	0.019
Mean Y	0.038	0.388	0.76	0.924	0.968

Panel D: number of children

	# Children (15-18)	# Children (19-25)	# Children (26-30)	# Children (31-35)	# Children (36-40)	# Children (41-49)
exposure (point)	−0.002	−0.037	−0.067	−0.151**	−0.038	−0.038
	(0.003)	(0.024)	(0.053)	(0.070)	(0.067)	(0.073)
Num.Obs.	20 747	56 848	32 510	26 469	24 899	35 064
R2 Adj.	0.011	0.306	0.198	0.232	0.254	0.282
Mean Y	0.006	0.624	1.882	2.429	2.735	2.993

Panel E: autonomy of women

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-49)
exposure (point)	0.035**	0.092***	0.022	0.006	0.070*
	(0.016)	(0.030)	(0.015)	(0.022)	(0.041)
Num.Obs.	24 411	5484	9572	7212	2143
R2 Adj.	0.148	0.138	0.097	0.091	0.106
Mean Y	0.635	0.505	0.639	0.704	0.718

Table E.4: Main results with point exposure and distance (i.e. without jittering correction)

Note: The tables repeat the paper's main regressions with point exposure as the treatment variable. Further, expected distance controls are replaced by point distance controls. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: years of education

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-19)	Higher Education (19-30)	All (5-30)
is female = 1 x exposure (point)	0.069*	0.162*	0.233*	0.178	0.309*	0.200**
	(0.040)	(0.090)	(0.140)	(0.134)	(0.158)	(0.102)
is female = 0 x exposure (point)	0.039	0.131*	0.042	0.102	0.062	0.089
	(0.036)	(0.074)	(0.099)	(0.124)	(0.161)	(0.100)
Num.Obs.	91 341	62 587	31 705	45 395	174 402	392 353
R2 Adj.	0.637	0.345	0.195	0.186	0.233	0.534
Mean Y	1.68	5.941	8.345	9.66	9.458	6.996

Panel B: degree obtained

	Primary	Secondary	Higher
is female = 1 x exposure (point)	0.021*	0.023	0.021
	(0.011)	(0.018)	(0.013)
is female = 0 x exposure (point)	0.004	0.010	0.011
	(0.010)	(0.018)	(0.013)
Num.Obs.	238 425	191 899	191 899
R2 Adj.	0.161	0.160	0.134
Mean Y	0.807	0.41	0.255

Panel C: is married

	Married (13-18)	Married (19-24)	Married (25-29)	Married (30-34)	Married (35-39)
is female = 1 x exposure (point)	−0.007	−0.028*	−0.013	0.002	0.005
	(0.007)	(0.016)	(0.022)	(0.008)	(0.005)
is female = 0 x exposure (point)	0.013***	0.003	−0.029	−0.019**	0.001
	(0.005)	(0.012)	(0.018)	(0.010)	(0.008)
Num.Obs.	95 359	87 467	68 256	57 081	46 469
R2 Adj.	0.060	0.284	0.193	0.069	0.019
Mean Y	0.038	0.388	0.76	0.924	0.968

Panel D: number of children

	# Children (15-18)	# Children (19-25)	# Children (26-30)	# Children (31-35)	# Children (36-40)	# Children (41-49)
exposure (point)	−0.002	−0.042*	−0.075	−0.145**	−0.036	−0.030
	(0.003)	(0.024)	(0.059)	(0.073)	(0.067)	(0.077)
Num.Obs.	20 747	56 848	32 510	26 469	24 899	35 064
R2 Adj.	0.011	0.306	0.198	0.232	0.254	0.282
Mean Y	0.006	0.624	1.882	2.429	2.735	2.993

Panel E: autonomy of women

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-49)
exposure (point)	0.037**	0.096***	0.021	0.009	0.070
	(0.016)	(0.029)	(0.015)	(0.022)	(0.043)
Num.Obs.	24 411	5484	9572	7212	2143
R2 Adj.	0.148	0.138	0.097	0.091	0.107
Mean Y	0.635	0.505	0.639	0.704	0.718

Table E.5: Observations at a distance of 50km from a radio station with point exposure and expected distance (i.e. only distance is corrected for jittering)

Note: The tables repeat the paper's main regressions with point exposure as the treatment variable. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

F Placebo and Robustness

	Primary	Primary	Secondary	Secondary	Higher	Higher
exposure	0.002		0.008		0.009	
	(0.015)		(0.017)		(0.014)	
Female x exposure		0.026		0.017		0.009
		(0.017)		(0.018)		(0.014)
Male x exposure		−0.021		−0.002		0.010
		(0.017)		(0.019)		(0.015)
Num.Obs.	108 266	108 266	108 266	108 266	108 266	108 266
R2 Adj.	0.225	0.225	0.152	0.152	0.120	0.120
Mean Y	0.636	0.636	0.253	0.253	0.155	0.155
Mean Y: Female	0.533	0.533	0.206	0.206	0.128	0.128
Mean Y: Male	0.74	0.74	0.299	0.299	0.183	0.183
SD Y	0.481	0.481	0.435	0.435	0.362	0.362
Controls	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓

Table F.1: Robustness: effect of exposure on education levels of individuals aged 30 to 40

Note: The tables regress the degree obtained for individuals aged 30 to 40 on exposure to radio. This age group is unlikely to actually be affected by radio in their educational choices, as the first radios launched when they were around 20 to 30 years old. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

	Lower Primary (30-35)	Upper Primary (36-39)	Lower Secondary (40-41)	Higher Secondary (42-44)	Higher Education (45-49)	All (30-49)
Female x exposure	0.172	0.329	0.515	−0.024	0.201	0.210
	(0.262)	(0.287)	(0.320)	(0.303)	(0.256)	(0.206)
Male x exposure	−0.070	−0.259	−0.114	−0.500	−0.049	−0.170
	(0.185)	(0.333)	(0.338)	(0.323)	(0.254)	(0.204)
Num.Obs.	75 194	34 978	22 838	22 935	44 097	200 042
R2 Adj.	0.263	0.283	0.286	0.307	0.315	0.294
Mean Y	7.872	7.472	6.588	6.875	5.974	7.123
Mean Y: Female	6.897	6.275	5.176	5.366	4.339	5.854
Mean Y: Male	8.841	8.771	7.93	8.464	7.608	8.406
SD Y	5.276	5.314	5.39	5.395	5.352	5.379
Controls	✓	✓	✓	✓	✓	✓
Radio FE	✓	✓	✓	✓	✓	✓

Table F.2: Robustness: effect of exposure on years of education of individuals aged 30 to 50

Note: The tables regress the years of education obtained for individuals aged 30 to 50 by age group on exposure to radio. These age groups are unlikely to actually be affected by radio in their educational choices, as the first radios launched when they were around 20 to 40 years old. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: Years of education

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-19)	Higher Education (19-30)	All (5-30)
Female x exposure	0.023	−0.075	−0.002	−0.054	−0.136	−0.080
	(0.043)	(0.099)	(0.108)	(0.180)	(0.167)	(0.107)
Male x exposure	0.033	−0.034	−0.038	−0.227	−0.291*	−0.118
	(0.045)	(0.078)	(0.138)	(0.169)	(0.157)	(0.089)
Num.Obs.	75 499	51 071	25 956	36 816	141 214	319 953
R2 Adj.	0.619	0.325	0.164	0.155	0.217	0.529
Mean Y	1.633	5.82	8.233	9.504	9.357	6.873
Mean Y: Female	1.657	5.874	8.234	9.368	8.755	6.686
Mean Y: Male	1.611	5.77	8.233	9.63	9.989	7.055
SD Y	1.618	2.146	2.574	3.407	4.869	4.821

Panel B: degree obtained

	Primary	Secondary	Higher
Female x exposure	−0.008	−0.026	−0.018
	(0.013)	(0.017)	(0.019)
Male x exposure	−0.018	−0.023*	−0.011
	(0.015)	(0.014)	(0.014)
Num.Obs.	155 525	155 525	155 525
R2 Adj.	0.155	0.148	0.124
Mean Y	0.785	0.397	0.24
Mean Y: Female	0.736	0.364	0.223
Mean Y: Male	0.836	0.431	0.258
SD Y	0.411	0.489	0.427

Panel C: is married

	Married (13-18)	Married (19-24)	Married (25-29)	Married (30-34)	Married (35-39)
Female x exposure	−0.011	0.014	0.000	0.006	0.004
	(0.007)	(0.020)	(0.026)	(0.015)	(0.012)
Male x exposure	0.008	0.034*	−0.008	−0.013	0.004
	(0.008)	(0.019)	(0.024)	(0.018)	(0.015)
Num.Obs.	77 620	70 702	55 367	46 686	37 032
R2 Adj.	0.068	0.271	0.181	0.066	0.027
Mean Y	0.042	0.392	0.754	0.916	0.962
Mean Y: Female	0.066	0.57	0.884	0.959	0.977
Mean Y: Male	0.019	0.204	0.616	0.873	0.947
SD Y	0.2	0.488	0.431	0.277	0.191

Panel D: number of children

	# Children (15-18)	# Children (19-25)	# Children (26-30)	# Children (31-35)	# Children (36-40)	# Children (41-49)
exposure	−0.002	0.028	0.009	0.117*	0.074	0.035
	(0.006)	(0.024)	(0.054)	(0.061)	(0.063)	(0.083)
Num.Obs.	16 891	46 318	26 299	21 207	19 935	27 673
R2 Adj.	0.015	0.295	0.204	0.221	0.231	0.257
Mean Y	0.008	0.623	1.85	2.426	2.753	3.055
SD Y	0.094	0.906	1.22	1.326	1.465	1.644

Panel E: autonomy of women

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-49)
exposure	−0.017	0.026	−0.024	−0.026	−0.055
	(0.025)	(0.036)	(0.035)	(0.034)	(0.056)
Num.Obs.	19 045	4364	7432	5607	1642
R2 Adj.	0.164	0.155	0.103	0.113	0.098
Mean Y	0.642	0.505	0.65	0.712	0.726
SD Y	0.326	0.332	0.316	0.306	0.301

Table F.3: Placebo linear model – observations at a distance of 50km from a radio station launched post 2015

Note: The tables repeat the paper's main regressions on the placebo sample, i.e., individuals in the vicinity and (potentially) coverage area of radios that launch post data collection (post 2015). Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Panel A: Years of education

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-19)	Higher Education (19-30)	All (5-30)
Female x exposure2	0.016	−0.097	−0.029	−0.240	0.053	−0.045
	(0.055)	(0.090)	(0.157)	(0.301)	(0.566)	(0.312)
Male x exposure2	0.012	−0.176***	−0.349*	−0.419**	−0.225	−0.232
	(0.028)	(0.053)	(0.202)	(0.192)	(0.240)	(0.155)
Num.Obs.	75 499	51 071	25 956	36 816	141 214	319 953
R2 Adj.	0.619	0.325	0.164	0.156	0.217	0.529
Mean Y	1.633	5.82	8.233	9.504	9.357	6.873
Mean Y: Female	1.657	5.874	8.234	9.368	8.755	6.686
Mean Y: Male	1.611	5.77	8.233	9.63	9.989	7.055
SD Y	1.618	2.146	2.574	3.407	4.869	4.821

Panel B: degree obtained

	Primary	Secondary	Higher
Female x exposure2	0.000	−0.015	−0.004
	(0.034)	(0.047)	(0.038)
Male x exposure2	−0.013	−0.039	−0.024
	(0.015)	(0.026)	(0.023)
Num.Obs.	155 525	155 525	155 525
R2 Adj.	0.155	0.148	0.124
Mean Y	0.785	0.397	0.24
Mean Y: Female	0.736	0.364	0.223
Mean Y: Male	0.836	0.431	0.258
SD Y	0.411	0.489	0.427

Panel C: is married

	Married (13-18)	Married (19-24)	Married (25-29)	Married (30-34)	Married (35-39)
is female = 1 x exposure2	−0.026*	−0.023	0.012	−0.009	−0.013
	(0.015)	(0.069)	(0.036)	(0.020)	(0.014)
is female = 0 x exposure2	−0.005	0.004	−0.082	−0.052	0.002
	(0.011)	(0.047)	(0.062)	(0.037)	(0.014)
Num.Obs.	77 640	70 719	55 370	46 710	37 048
R2 Adj.	0.068	0.271	0.181	0.066	0.027
Mean Y	0.042	0.392	0.754	0.916	0.962
SD Y	0.2	0.488	0.431	0.277	0.19

Panel D: number of children

	# Children (15-18)	# Children (19-25)	# Children (26-30)	# Children (31-35)	# Children (36-40)	# Children (41-49)
exposure2	−0.005	−0.010	−0.083	−0.035	0.105	0.082
	(0.003)	(0.043)	(0.070)	(0.096)	(0.162)	(0.171)
Num.Obs.	16 891	46 318	26 299	21 207	19 935	27 673
R2 Adj.	0.015	0.295	0.204	0.221	0.231	0.257
Mean Y	0.008	0.623	1.85	2.426	2.753	3.055
SD Y	0.094	0.906	1.22	1.326	1.465	1.644

Panel E: autonomy of women

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-49)
exposure2	−0.009	0.047**	−0.064*	0.013	0.052
	(0.026)	(0.022)	(0.035)	(0.035)	(0.048)
Num.Obs.	19 045	4364	7432	5607	1642
R2 Adj.	0.163	0.156	0.103	0.113	0.097
Mean Y	0.642	0.505	0.65	0.712	0.726
SD Y	0.326	0.332	0.316	0.306	0.301

Table F.4: Placebo quadratic model – observations at a distance of 50km from a radio station launched post 2015

Note: The tables repeat the paper's main regressions on the placebo sample, i.e., individuals in the vicinity and (potentially) coverage area of radios that launch post data collection (post 2015). Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

G Alternative Specifications

Panel A: years of education

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-19)	Higher Education (19-30)	All (5-30)
is female = 1 x exposure	0.070	0.245**	0.392***	0.282	0.493**	0.309**
	(0.046)	(0.098)	(0.146)	(0.200)	(0.241)	(0.143)
is female = 0 x exposure	0.051	0.223**	0.139	0.121	0.195	0.178
	(0.046)	(0.099)	(0.147)	(0.187)	(0.246)	(0.146)
Num.Obs.	91 341	62 587	31 705	45 395	174 402	392 353
R2 Adj.	0.637	0.345	0.195	0.186	0.233	0.534
Mean Y	1.68	5.941	8.345	9.66	9.458	6.996
SD Y	1.635	2.116	2.584	3.406	4.912	4.855

Panel B: degree obtained

	Primary	Secondary	Higher
is female = 1 x exposure	0.034**	0.037*	0.027
	(0.017)	(0.022)	(0.018)
is female = 0 x exposure	0.010	0.017	0.015
	(0.017)	(0.023)	(0.019)
Num.Obs.	191 899	191 899	191 899
R2 Adj.	0.166	0.160	0.134
Mean Y	0.788	0.41	0.255
SD Y	0.409	0.492	0.436

Panel C: is married

	Married (13-18)	Married (19-24)	Married (25-29)	Married (30-34)	Married (35-39)
is female = 1 x exposure	−0.015**	−0.051***	−0.025	0.009	0.005
	(0.007)	(0.018)	(0.016)	(0.011)	(0.006)
is female = 0 x exposure	0.009	−0.015	−0.042**	−0.017	−0.004
	(0.006)	(0.018)	(0.019)	(0.013)	(0.008)
Num.Obs.	95 359	87 467	68 256	57 081	46 469
R2 Adj.	0.060	0.284	0.193	0.069	0.019
Mean Y	0.038	0.388	0.76	0.924	0.968
SD Y	0.192	0.487	0.427	0.265	0.175

Panel D: number of children

	# Children (15-18)	# Children (19-25)	# Children (26-30)	# Children (31-35)	# Children (36-40)	# Children (41-49)
exposure	−0.001	−0.079**	−0.138**	−0.210***	−0.033	−0.012
	(0.003)	(0.031)	(0.063)	(0.077)	(0.074)	(0.082)
Num.Obs.	20 747	56 848	32 510	26 469	24 899	35 064
R2 Adj.	0.011	0.306	0.198	0.232	0.254	0.282
Mean Y	0.006	0.624	1.882	2.429	2.735	2.993
SD Y	0.084	0.899	1.199	1.295	1.43	1.633

Panel E: autonomy of women

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-54)
exposure	0.037	0.120***	0.009	−0.001	0.068
	(0.023)	(0.041)	(0.030)	(0.029)	(0.053)
Num.Obs.	24 411	5484	9572	7212	2143
R2 Adj.	0.147	0.138	0.097	0.091	0.106
Mean Y	0.635	0.505	0.639	0.704	0.718
SD Y	0.329	0.336	0.322	0.309	0.308

Table G.1: Main results with linear treatment effect over time: observations at a distance of 50km from a radio station

Note: The tables repeat the paper's main regressions using clustered standard errors. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Heteroscedasticity robust standard errors clustered at the subdistrict level reported in parentheses.

Panel A: years of education

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-19)	Higher Education (19-30)	All (5-30)
is female = 1 x exposure2	0.062	0.252**	0.494***	0.522**	0.770***	0.461***
	(0.049)	(0.110)	(0.152)	(0.235)	(0.265)	(0.157)
is female = 0 x exposure2	0.053	0.259**	0.137	0.418**	0.552**	0.370**
	(0.046)	(0.108)	(0.159)	(0.213)	(0.277)	(0.166)
Num.Obs.	91 341	62 587	31 705	45 395	174 402	392 353
R2 Adj.	0.637	0.345	0.195	0.186	0.233	0.534
Mean Y	1.68	5.941	8.345	9.66	9.458	6.996
SD Y	1.635	2.116	2.584	3.406	4.912	4.855

Panel B: degree obtained

	Primary	Secondary	Higher
is female = 1 x exposure2	0.054***	0.063***	0.044**
	(0.018)	(0.024)	(0.019)
is female = 0 x exposure2	0.029	0.059**	0.052**
	(0.019)	(0.025)	(0.021)
Num.Obs.	191 899	191 899	191 899
R2 Adj.	0.166	0.160	0.134
Mean Y	0.788	0.41	0.255
SD Y	0.409	0.492	0.436

Panel C: is married

	Married (13-18)	Married (19-24)	Married (25-29)	Married (30-34)	Married (35-39)
is female = 1 x exposure2	−0.023***	−0.061***	−0.043**	0.010	−0.002
	(0.008)	(0.021)	(0.017)	(0.012)	(0.007)
is female = 0 x exposure2	0.009	−0.019	−0.064***	−0.018	−0.008
	(0.007)	(0.021)	(0.022)	(0.015)	(0.009)
Num.Obs.	95 359	87 467	68 256	57 081	46 469
R2 Adj.	0.060	0.284	0.193	0.068	0.019
Mean Y	0.038	0.388	0.76	0.924	0.968
SD Y	0.192	0.487	0.427	0.265	0.175

Panel D: number of children

	# Children (15-18)	# Children (19-25)	# Children (26-30)	# Children (31-35)	# Children (36-40)	# Children (41-49)
exposure2	−0.001	−0.097***	−0.188***	−0.306***	−0.164**	−0.141
	(0.004)	(0.032)	(0.066)	(0.076)	(0.079)	(0.086)
Num.Obs.	20 747	56 848	32 510	26 469	24 899	35 064
R2 Adj.	0.011	0.306	0.199	0.233	0.255	0.282
Mean Y	0.006	0.624	1.882	2.429	2.735	2.993
SD Y	0.084	0.899	1.199	1.295	1.43	1.633

Panel E: autonomy of women

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-54)
exposure2	0.018	0.104**	−0.003	−0.028	0.065
	(0.026)	(0.049)	(0.034)	(0.030)	(0.057)
Num.Obs.	24 411	5484	9572	7212	2143
R2 Adj.	0.147	0.137	0.097	0.091	0.106
Mean Y	0.635	0.505	0.639	0.704	0.718
SD Y	0.329	0.336	0.322	0.309	0.308

Table G.2: Main results with quadratic treatment effect over time: observations at a distance of 50km from a radio station

Note: The tables repeat the paper's main regressions using clustered standard errors. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Heteroscedasticity robust standard errors clustered at the subdistrict level reported in parentheses.

Panel A: years of education

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-19)	Higher Education (19-30)	All (5-30)
is female = 1 x exposure	0.106*	0.319***	0.426**	0.266	0.561**	0.372***
	(0.065)	(0.099)	(0.172)	(0.192)	(0.239)	(0.135)
is female = 0 x exposure	0.109*	0.261***	0.195	0.131	0.205	0.204
	(0.057)	(0.084)	(0.121)	(0.214)	(0.223)	(0.132)
Num.Obs.	69 057	47 426	24 042	34 802	135 928	301 132
R2 Adj.	0.644	0.350	0.195	0.192	0.234	0.540
Mean Y	1.69	5.988	8.408	9.724	9.608	7.118
SD Y	1.641	2.099	2.555	3.392	4.892	4.884

Panel B: degree obtained

	Primary	Secondary	Higher
is female = 1 x exposure	0.040**	0.040	0.028
	(0.016)	(0.026)	(0.018)
is female = 0 x exposure	0.019	0.012	0.011
	(0.014)	(0.025)	(0.018)
Num.Obs.	184 649	149 339	149 339
R2 Adj.	0.163	0.162	0.137
Mean Y	0.815	0.422	0.266
SD Y	0.389	0.494	0.442

Panel C: is married

	Married (13-18)	Married (19-24)	Married (25-29)	Married (30-34)	Married (35-39)
is female = 1 x exposure	−0.015**	−0.054**	−0.026	0.006	0.000
	(0.007)	(0.024)	(0.027)	(0.014)	(0.009)
is female = 0 x exposure	0.012**	−0.015	−0.045*	−0.016	−0.003
	(0.005)	(0.019)	(0.024)	(0.017)	(0.011)
Num.Obs.	72 505	67 785	53 550	44 601	36 159
R2 Adj.	0.058	0.281	0.194	0.069	0.021
Mean Y	0.037	0.378	0.753	0.921	0.967
SD Y	0.19	0.485	0.431	0.269	0.179

Panel D: number of children

	# Children (15-18)	# Children (19-25)	# Children (26-30)	# Children (31-35)	# Children (36-40)	# Children (41-49)
exposure	0.000	−0.081*	−0.153*	−0.243**	0.018	−0.044
	(0.004)	(0.043)	(0.079)	(0.108)	(0.097)	(0.114)
Num.Obs.	15 668	43 661	25 299	20 403	19 369	27 033
R2 Adj.	0.010	0.301	0.204	0.235	0.257	0.286
Mean Y	0.006	0.612	1.852	2.39	2.683	2.937
SD Y	0.086	0.893	1.197	1.28	1.4	1.609

Panel E: autonomy of women

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-49)
exposure	0.040*	0.125***	0.004	0.009	0.081
	(0.021)	(0.040)	(0.023)	(0.033)	(0.072)
Num.Obs.	18 583	4116	7315	5514	1638
R2 Adj.	0.156	0.142	0.107	0.096	0.109
Mean Y	0.64	0.505	0.643	0.711	0.722
SD Y	0.328	0.336	0.321	0.306	0.306

Table G.3: Observations at a distance of 40km from a radio station

Note: The tables repeat the paper's main regressions reducing the sample distance to 40km. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter V.1. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

H Data on Radios: Data Gathering and Preparation

Data on Community Radios and their locations was manually gathered from a number of sources.

The precise information on the location was then hand collected from various sources in May 2021. The starting point was always the list of radio stations issued by the Ministry of Broadcasting and Information in March 2020 (MOIB, 2020). This list includes adresses, names of organizations and other information. Furthermore, this information is enriched using a list of operational stations compiled by Jacob (2021) of the National Institute of Amateur Radio and other sources, mainly from MOIB (e.g. CRFC, 2021).[27] Based on this data, radios are searched for via Google Maps. Many of these stations have their own Google Maps entry and were geocoded accordingly. Others are identified via their parent organization. Locations are verified (where possible) via websites and by searching for pictures taken and posted on Google Maps in the vicinity of a radio tower (e.g., of local stores). In total, 276 out of 289 stations in the list were identified as operational as of May 2020. Of these, 264 or 96% could be precisely geocoded using the above approach. In the process, I identified 110 radio towers in pictures, which verifies the precision of the location. Finally, the MOIB shared a list of radio tower coordinates with me. Unfortunately, this list only had a precision of around 1.2km. However, I used it to verify and improve the precision of coordinates in my data.

Regarding technical specifications, radios are, by regulation, limited to transmitting at a power of 50W, putting Indian CRS at the lower end of the typical power permitted to a CRS (Fraser and Estrada, 2001), and to building towers at a height of 30m (Govt. of India, 2006). Based on multiple interviews with experts, NGOs working with CRS and MOIB, as well as visits to multiple community radio stations and receiving reports on visits from Jose Jacob at the National Institute of Amateur Radio, I verified that virtually all radios maximize their coverage by transmitting at this frequency using 30m or close to 30m towers.

I Radio Content I: Information from Compendia

In total, I collected information on radios' self-descriptions. The primary source of this are Radio Compendia. These are regularly created booklets that summarize the content of a given radio station on a single page, as shown in Figure I.1. They are created as part of the Community Radio Sammelan, a facilitation event for CRS. In case such data is not available, websites of radios and other sources on the radio (such as articles in newspapers) are searched for information on the content. In total, I collect information on the content of 237 out of 264 CRS or around 90%. 90% of the content information stems from Compendia, most of which is from the 2019 edition (180 of 211).

The content information is then manually coded by topic. First, I go through the compendia to identify the main topics mentioned. Next, I use CATMA (Gius et al., 2023), a QTA text annotation software, to manually annotate words that are related to the respective topic. I tag texts in two categories: words related to content and words related to a radio's audience, format, or protagonists. In the coding process I follow the following logic:

For content related words, I only tag words that directly relate to the respective topic and are required to understand the context. This usually does not include the entire sentence. For example: *"The radio is focused on women empowerment, in particular child marriage and dowry"*. Words that are ambiguous with respect to whether they relate to a given topic are only marked if the text contains other words that make this link clear. For example "skill development" is only marked under "economic" if the text contains a word related to the economic development of listeners, such as "career guidance".

The following topics are coded:

- agriculture & fishing: e.g. advise and technology transfer
- culture: anything related to the preservation of local culture, such as the support of local talent
- economic: specifically focuses on furthering individuals' economic well-being, e.g., entrepreneurship, personal finance, career counselling etc. (excl. agricultural adivse)
- education: e.g. educational programs or underlining the importance of education
- environment: environmental concerns and disaster prevention and mitigation

[27] Thank you to Jose Jacob for also sharing his reports from visits to multiple CRS with me, including on their technical details.

- governance: local governance and information on government schemes
- social empowerment & rights: focus on the legal rights of marginalized groups and the empowerment of marginalized groups such as ST/SC (excl. women and children, except if legal rights of these groups were explicitly mentioned). Note that I did not include generic words such as 'social issues' or 'social development', as these are ambiguous.
- women empowerment: topics related to the empowerment of women, e.g., dowry, child marriage, girls' education etc.
- health & hygiene: focus on health information, including nutrition, disease information (TB, HIV/AIDS, etc.), sanitation, and hygiene
- youth empowerment: focus on empowering youth, including children and adolescents specifically.

Further, regarding radios' format, content, and audience, I further marked every word related to these topics.

ANDHRA PRADESH

Radio Vishnu

90.4 MHz

Radio Vishnu

Shri Vishnu Engineering College for Women,
Vishnupur, Bhimavaram. Andhra Pradesh - 534 202
Website: www.radiovishnu.com
Contact: Dr. S. Hanumantha Rao
Email: radiovishnu@gmail.com
Phone: 09849782622, 08816250864

Launch Date
15-4-2007

Broadcast Timing
6.00 AM - 10.00 AM
12.00 PM - 20.00 PM

Broadcast Hours
12 hours

Languages
Telugu, Hindi and English

Enabling students to acquire skills

Photo Courtesy: Radio Vishnu

GENESIS
The management of Sri Vishnu Educational Society realised the crucial role of community radio in uplifting village people socially. The station was set up to provide students with ample opportunities to acquaint themselves practically with radio and TV technologies. In this way, the station was created to enable students to acquire technical management skills for better career opportunities.

THEMATIC FOCUS
The community radio station station focusses on various issues, including health and nutrition, women empowerment and agriculture. The station intends to work around themes which are directly related to the local community.

SIGNATURE PROGRAMMES
Rythanna, Aharam Arogyam, Mahila and Balavinodini.

INNOVATIONS IN FORMAT
Expert talks and interviews are given by Krishi Vigyan Kendra (KVK) scientists. Also, many outreach programmes are undertaken in different villages for radio broadcast. The station also broadcasts the experiences of farmers to highlight their problems.

KEY PARTNERS
Krishi Vigyan Kendra (KVK), Undi; Rice Research Station, Maruteru; Vishnu Dental College, Bhimavaram

CORE TEAM
V.V. Subrahmanyam; T. Madhu; K. Vijaya Lakshmi and P.N.V. Krishna

CRMC MEMBERS
G. Srinivasa Rao, J. Prasad Raju, P. Srinivasa Raju, Dr. U. V. Ramana Raju, P. Annamani, Dr. S. Hanumantha Rao.

5

Figure I.1: Example page of radio compendium of 'Radio Vishnu'

In a next step, the words and information on the related radio station are exported. To get an idea on the distribution of topics over radios, I first create a dummy for each radio indicating whether it mentions each of the above topics as one of their key themes (see Figure I.3). To get a better idea of the content of each topic, I further create wordclouds as shown in Figure I.2. These are created after further pre-processing the highlighted words by removing stop words, moving to lower case, and steming and by removing infrequent terms. Finally, Figure I.4 gives an idea of the formats in which programs are produced, who appears on radio, and who listens to radio according to radio compendia.

J Radio Content II: Information from Radio Show Recordings

J.1 Data Preparation and Topic Model

This subsection describes the application of the topic model and its underlying data regarding radios' audio files.

Starting with Table J.1, the underlying data from `edaa.in` is discussed. It shows the number of shows uploaded by each radio station. As visible, a couple of radio stations are responsible for most of the content. Regarding the format, CRS indicate that shows are produced in a variety of formats as Table J.2, ranging from discussions to documentaries, music, and phone-in/-out shows. Similarly, these are heterogeneous with respect to the languages in which content is produced: Table J.3 shows that content is uploaded in 22 languages.

To obtain a better idea of the shows' content, I next draw on the audio files of uploads to `edaa.in` to transcribe the shows. More specifically, I use *Google's Speech-to-Text API* to transcribe shows in supported languages. Supported languages account for 92% of shows uploaded and cover 105 of 114 radios.[28] I transcribe up to 586 shows per radio. Table J.1 shows that only four radios upload more content. For these, I randomly choose 586 shows. Next, *Google Translate* is used to translate transcripts into English. A total of 6,509 shows are transcribed and translated. The average show has a length of 12:30 min. Some of the shows are uploaded twice. After removing duplicate transcripts (597 shows), non-English ones, i.e. where translation or transcription failed - 68 shows), and exceptionally long (>10k tokens - 28 shows) or short (<20 tokens - 59) transcripts, I end up with a total of 5,806 shows produced by 93 stations (85% of stations that uploaded content).

Next, the transcripts are pre-processed by removing punctuation, non-English characters, and stop words; changing characters to lower case ones and steming words. Terms that appear in less than 10 documents or less than 0.1% of transcripts are removed. Next, I calculate the Term Frequency-Inverse Document Frequency (tf-idf) matrix. This adjusts the Term Frequency (i.e., the number of times a term appears in a document) by the logarithm of the inverse of the share of documents a term appears in. In effect, this gives more weight to terms unique to specific types of documents, i.e., potentially informative about their topic. At the same time, it punishes terms that appear across most documents. Figure J.1 shows the 150 terms with the highest tf-idf. As shown, the term 'women' has the highest weight, i.e., appearing both often but only in specific documents. Following Grün and Hornik (2011), I use the tf-idf to reduce the Document Frequency Matrix (dfm) to terms that are relevant. In particular, I rank the terms by tf-idf and remove those with a rank below 8,000.

[28] Available: Telugu: te-IN; English: en-IN; Hindi: hi-IN; Urdu: ur-IN; Malayalam: ml-IN; Gujarati: gu-IN; Kannada: kn-IN; Punjabi: pa-Guru-IN; Marathi: mr-IN; Tamil: ta-IN. Missing Languages: Assamese, Bangala, Bhojpuri, Khasi, Bundeli, Surgujiha, Mev, Maithili, Oriya, Rajasthani, Garhwali.

Figure J.1: Wordcloud of document frequency (TF-IDF) matrix: 150 terms with highest weight

Note: The wordcloud is based on translated transcripts of radio shows uploaded to `edaa.in`.

Term frequencies are then used to predict 35 discussion topics using Latent Dirichlet allocation (LDA) (Blei et al., 2003). Each document is assigned a topic share. 35 topics were chosen based on multiple trial runs. A lower number created additional topics that appeared very broad. A higher number did not add much nuance. LDA is arguably the most widely used method to determine latent topics in a selection of documents. Intuitively, it treats each transcript as a mix of latent topics, where topics are probability distributions over words. A combination of topics thus characterizes documents. A document's content is then created by sampling words from those topics. Based on these assumptions, the terms, and the chosen number of topics, LDA aims to estimate the topic distribution for each document and the term distribution for each topic (Hansen et al., 2018, provide a detailed description of LDA, including its underlying econometrics).[29]

The topics are hand-labeled, building on each topic's 15 most predictive terms. These can be viewed in Table J.4. To get an idea of CRS' content, I first collapse the topics into 11 broader main topics, ranging from health to education and women-specific topics. Figure 2 visualizes different radios' shares across these. As is visible, radios are quite heterogeneous regarding their content. Some radios strongly focus on women-related issues.

To gain a deeper understanding of content targeted toward women, Figure J.2 concentrates on shows that specifically target this demographic. In particular, I identify topics specifically targeted toward women. These typically include the term 'women', 'girl' or a synonym in the top 15 words and concern issues specifically relevant to women/girls, such as (child) marriage, pregnancy, menstruation etc. The Figure shows the proportion of various categories with women-centric content produced by different radio stations. The radios are shown in the same order as in Part I. As is visible, CRS with the highest intensity of women-related programs produce a higher share of shows regarding the social role of women, as indicated by the high share of the 'marriage' category. The other topics appear more evenly distributed across radios.

[29] Operations are done using the Quanteda Package in R (Benoit et al., 2018). LDA is applied using the seededlda package in R (Watanabe and Xuan-Hieu (GibbsLDA++), 2023)

Radio	Shows Uploaded	Radio.1	Shows Uploaded
Aap Ki Awaaz	4	Periyar CR	1
Aapno Radio	4	PGP Radio	1
AGN CRS	87	PSG CR	4
Agra Ki Awaaz	1	Puduvai Vani	49
Alfaz-e-Mewat	422	Radio 7	2
Alwar ki Awaz 90.8MHz	64	Radio Active CR 90.4	94
Anna Community Radio	201	Radio Adan	2
Apna Radio	50	Radio Ala 90.8	8
BBD 90.8 FM	4	Radio Azad Hind	128
Bol Hyderabad	5	Radio Benziger	2,658
Brahmaputra Community Radio Station	1	Radio Bundelkhand	4
Chanderi Ki Awaaz	106	Radio Dhadkan 107.8 MHz	32
Chitkara	23	Radio Eminent	1
CMS Radio Lucknow	25	Radio FTII	46
CMS RADIO LUCKNOW	39	RADIO JAGRITI	1
Deccan Radio	37	Radio Jamia	2
Divya Vani Neladani	1	Radio JU	134
ENTE RADIO	2	Radio Khushi	44
GNGC CR	1	Radio Luit	494
Green Radio	61	Radio Macfast	1
Gurgaon Ki Awaaz Samudayik Radio Station	79	Radio Madhuban	154
Guruvani	23	Radio Mahananda 98.8 FM	10
Hello Doon	101	Radio Manav Rachna	78
Hello Haldwani	6	Radio Mangalam	26
Himgiri ki awaaz	7	Radio Manipal	1
HINT CR	2	Radio Mattoli	586
Holy Cross CR	6	Radio Media Village	1,538
Honey CR	6	Radio Mewat	278
IIT CR	114	Radio Nagar 90.4 FM - Awaj Tumcha	2
Janadhwani	3	Radio Namaskar	351
JIMS Radio	2	Radio Popcorn	1
Jnan Taranga	21	Radio Rimjhim	3,787
Jyotirgamaya CR	8	Radio Sirsa	1
Kalanjiam Samuga	9	Radio Snehi	90
Kalpakkam CRS	25	Radio SRFTI	3
Kamalvani	44	Radio Vishwas 90.8	1
Kisan Vani	23	Rudi No Radio	37
KMIT Tarang	1	Samudayik Radio Henvalvani	2
Kongu CR	45	Sanjha Radio 90.8 MHz	1
Krishi CRS	49	Sarang CR	1
KSR Community Radio	5	sarathi jhalak	66
Kumaon Vani	6	Sharda Krishi Vahini	1
KVK Pravara CR	1	Shyamalavani	352
Lalit Lokvani	38	SSM CR	1
Manndeshi Tarang Vahini	8	Styavani	4
MOP CR	1	Suno Sharda	1
MSPICM CR	278	Thendral CR	1
Mukta Vidya Vani	93	Tilonia Radio	36
MUST Radio	101	Vasundhara Krishi Vahini	56
Muthucharam CR	77	Vayalaga vanoli	123
Namma Dhwani	4	VENUDHWANI KLE KANASU 90.4	2
NAV JAGRITI YUVA MANDAL	1	Vidyavani Community Radio	62
Neotech CR	23	VIT Community Radio 90.8	60
Nila CR	1	Vivek CR	1
Pantnagar Janvani	181	Voice of SOA Community	2
PARD Vanoli	199	Y-FM	24
pasumaifm	2	Yeralavani	2

Table J.1: Number of shows uploaded by radio station

Note: The above table describes the number of shows uploaded by each radio station to edaa.in.

Format	Num. Shows	Num. Radios
Discussion	832	46
Documentry	1135	28
Drama	761	46
Feature	636	47
Interview	1702	61
Jingle	420	30
Magazine	2236	60
Music	805	49
News	13	9
Phone in/out	1482	18
Radio Spot	247	28
Talk	3747	58
Vox-Populi	163	13

Table J.2: Number of shows uploaded by type of radio show

Note: The above table describes the number of shows uploaded under in respective format on `edaa.in`. It further shows the number of radios that uploaded any show to in the respective format.

Language	Num. Shows	Num. Radios
Assamese	469	3
Bangala	187	6
Bhojpuri	871	2
Bundeli	35	4
English	228	20
Garhwali	2	1
Gujarati	64	3
Hindi	5359	63
Kannada	216	11
Khasi	3	1
Kumaoni	1	1
Maithili	2	2
Malayalam	4811	8
Marathi	233	12
Mev	32	2
Oriya	353	4
Punjabi	4	3
Rajasthani	36	3
Surgujiha	2	1
Tamil	1224	22
Telugu	17	5
Urdu	30	3

Table J.3: Number of shows uploaded by language of radio show

Note: The above table describes the number of shows uploaded under in respective language on `edaa.in`. It further shows the number of radios that uploaded any show to in the respective language. Available languages on Google's Speech-to-Text API: Telugu, English, Hindi, Urdu, Malayalam, Gujarati, Kannada, Punjabi, Marathi, Tamil. Missing Languages: Assamese, Bangala, Bhojpuri, Khasi, Bundeli, Surgujiha, Mev, Maithili, Oriya, Rajasthani, Garhwali.

agriculture	communityfamily	edu_general1	edu_general2	edu_general3	edu_general4	edu_math1
farmer	brother	children	work	know	yes	equal
crop	hous	school	today	tell	school	x
plant	mother	teacher	peopl	like	tell	squar
agricultur	know	child	educ	look	time	triangl
soil	yes	time	good	answer	today	angl
water	take	parent	student	yes	answer	b
farm	tell	studi	india	time	number	point
seed	work	educ	develop	good	good	geometri
cow	happen	father	govern	everyon	know	minus
product	son	tell	train	work	two	line
fertil	today	mother	import	long	equal	one
land	day	friend	institut	learn	take	area
cultiv	daughter	thing	person	ali	question	take
day	look	want	countri	new	just	mathemat
kg	right	read	area	find	one	y

edu_math2	edu_math3	entertainment_festivals	entertainment_india1	entertainment_india2	entertainment_music1	entertainment_music2
good	mathemat	day	swamiji	india	song	re
day	number	countri	time	year	love	raga
peopl	book	india	vivekananda	first	heart	sa
littl	one	name	swami	team	friend	ga
lot	scienc	celebr	peopl	indian	listen	music
one	video	gandhi	india	world	life	ma
mani	like	peopl	day	time	film	ra
know	comput	holi	ji	film	beauti	song
time	mathematician	prophet	religion	countri	eye	pa
next	interest	king	countri	run	like	tell
rupe	program	british	even	minist	voic	jai
two	technolog	peac	shri	new	name	program
question	call	world	start	cricket	world	dha
ask	differ	allah	mani	last	color	nana
thing	read	histori	god	second	happi	na

entertainment_quiz	entertainment_spiritual	environment	govt_banking	govt_programs1	health_general1	health_lung
question	life	water	bank	villag	eye	tv
answer	mind	environ	inform	group	bodi	diseas
start	world	tree	money	panchayat	blood	cancer
first	live	clean	govern	program	breath	hiv
option	think	pollut	give	work	donat	peopl
time	good	earth	given	gram	place	know
call	man	rain	road	sabha	yoga	patient
second	person	plant	take	sister	peopl	treatment
next	make	garbag	interest	sarpanch	nose	lung
name	god	save	peopl	meet	way	doctor
studi	one	place	land	hous	food	medicin
contest	happi	peopl	account	raj	gayatri	spread
readi	peopl	river	number	govern	diseas	smoke
correct	thing	drink	problem	everi	exercis	person
bihar	way	citi	offic	panchayati	hand	tuberculosi

health_nutrition	health1	other1	other2	other3	other4	other5
eat	diseas	hai	thing	ji	abl	ji
food	doctor	ki	tell	speak	good	peopl
veget	blood	mein	like	peopl	make	like
make	problem	ke	time	good	time	sudhakar
bodi	bodi	ka	much	rimjhim	know	tell
vitamin	medicin	aur	talk	issu	case	nowaday
milk	pain	se	peopl	thank	first	thing
diet	time	hain	one	call	day	time
good	patient	kya	take	thing	lot	one
fruit	due	math	mani	much	way	mani
take	caus	nahin	want	today	take	day
drink	take	ko	good	welcom	learn	start
green	reason	bhi	know	keep	peopl	year
women	like	liy	happen	understand	media	vinita
protein	stomach	per	lot	talk	world	lot

rights	women_edu	women_health	women_health1	women_health2	women_marriage	women_maternity
right	health	eat	women	health	girl	child
constitut	scienc	bodi	program	diseas	famili	mother
court	sister	blood	healthi	council	women	children
countri	technolog	iron	talk	nation	child	month
law	water	anemia	time	bodi	marriag	babi
peopl	diseas	girl	tell	scienc	woman	milk
talk	program	problem	like	abl	children	time
live	women	food	health	life	marri	day
program	hand	new	problem	technolog	year	care
think	talk	talk	take	treatment	daughter	take
govern	keep	mani	know	project	societi	give
life	take	increas	woman	like	age	pregnanc
ji	depart	hous	mani	women	boy	deliveri
work	tell	hemoglobin	much	make	husband	pregnant
thing	clean	tea	care	blood	program	vaccin

Table J.4: Top 15 most predictive words by topic of LDA topic model

Note: The above table shows the 15 most predictive words for each of the 35 topics of the LDA model. The first word in each list describes the topic and, if applicable, subtopic, separated by "_". In case multiple groups of words are categorized under the same topic or subtopic, the topic name includes a count.

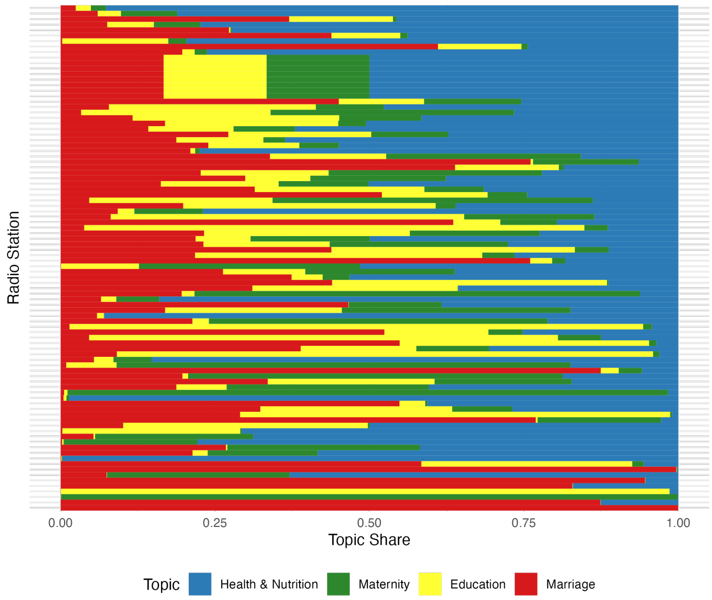


Figure J.2: Radios' share of content across women-related topics

Note: The above figure shows radio stations on the vertical axis. The horizontal axis shows the distribution of content in the "women" topic to the respective subtopics.

J.2 GPT-Based Content Analysis

Here, I provide additional details on the content analysis of transcripts. Unless otherwise specified, the temperature of the GPT request is set to 0. This makes GPT more likely to choose words with the highest probability, meaning that the results better replicate and reduced risk of 'halucination'. Further, I always set the model's role to "You are a helpful research assistant". This guides how the model will behave. In Rusche (2024), I provide a short summary on how LLMs can be called using R.

Starting with the preparation of transcripts, I first send all articles to ChatGPT-3.5 to restore grammatical structure while aiming to leave the content intact and without adding additional information. I choose GPT-3.5 because it has a substantially higher output response length than GPT-4o. The prompt is as follows:[30]

> **Prompt: Text Restauration**
>
> The following text is a translated transcript of an Indian radio show, which has lost its grammatical structure in translation. Please reconstruct the text to restore its original coherence and readability without adding any new content. Return only the revised text without any additional comments or preface: "[text]"

Next, I ask GPT-4 to return a vector for whether the respective show covers one of four topics of interest.

[30] Whenever the text was longer than the allowed context window for GPT-3.5, it was split in equal parts which were send separately.

Prompt: Topic of Text

'The following text is a translated transcript of an Indian radio show. Please answer the following four questions only with yes or no. The questions are:

1. Does this program cover the topic of child or early marriage?
2. Does this program cover the topic of education of girls?
3. Does this program cover any of the following topics: fertility, contraception, or family planning?
4. Does this program cover the topic of domestic violence or violence against women?
5. Describe the underlying topic of the program in at most 5 words.

The answer should only contain a vector with the answers: c("yes or no", "yes or no", "yes or no", "yes or no", "description") without any additional comments or preface. The text is: ' "[text]"

This returns a vector with four binary variables indicating whether a given topic was covered. In total, I send up to three requests per text. If the first two agree, i.e. return the same vector, I take this result; otherwise I send a third request and define the final vector via majority rule. Given that I, hence, want some diversity in the answers, I set the temperature to 0.1.[31]

Next, I additionally identify articles on the issues of interest via simple keyword search. Specifically, I define an article as covering a specific topic if it contains any of the following words:

- child marriage: "child marriage", "early marriage"
- girls' education: ("girl" or "girls" or "female") AND ("education" or "school")
- fertility: "sterilization", "condom", "condoms", "ovulation", "contraception", "contraceptive", "birth control", "family planning", "reproductive rights"
- violence against women: "violence", "intimate partner violence", "domestic abuse", "spousal abuse", "partner abuse", "family violence", "marital abuse", "intimate violence", "domestic conflict", "domestic maltreatment"

Finally, I then send the text to ChatGPT-4 again, this time asking it to return a list of answers. The first answer is about whether or not the respective topic is covered (a single request is send for every topic covered by the article). This is particularly relevant for shows identified as covering a topic via keywords. In case this question is answered with "yes", I ask two additional questions. The first is about whether the text is in favor, neutral or against a specific issue (e.g. child marriage). The final question then asks which arguments are put forward if a 'progressive' stance is taken.

Prompt: Position/Stance Taken by Radio Show

"The following text is a transcript of an Indian radio show. Fill the following list. In case the question does not apply, simply enter NA into the list: 'list("Does this program cover the topic of child or early marriage?" = "Yes or No", "If yes, is the programs message or plot in favor or against or neutral towards child/early marriage?" = "in favor/neutral/against", "If against, briefly summarize up to three arguments (may be less if less than 3 are mentioned) in bullet points that the program explicitly makes against child/early marriages." = c("Argument1", "Argument2", "Argument3"))' Return only the full list without any additional comments or preface. Here is the transcript: "[text]"

The result is a list that can be parsed into R. In rare cases where parsing the returned object fails, the request is sent again.

[31] I also build in quality checks. Following every request, these check whether the answer returned is a vector. If not, the request is send again.

(a) Governance

(b) Health & Hygiene

(c) Education

(d) Agriculture/Fishing

(e) Culture

(f) Youth Empowerment

(g) Environment

(h) Economic

(i) Women Empower.

(j) Social Empowerment & Rights

Figure I.2: Word clouds by topic of radio stations

Note: The wordcloud is created by pooling all words or sentences that were manually coded to be related to the respective topic. These are pre-processed by removing stop words, changed to lower case letters, and stemmed. Next, 1, 2, and 3 grams are created and used to plot the wordcloud based on term frequencies. Words are scaled by the square root of their frequency as per default in the *ggwordcloud* package (Le Pennec and Slowikowski, 2023).

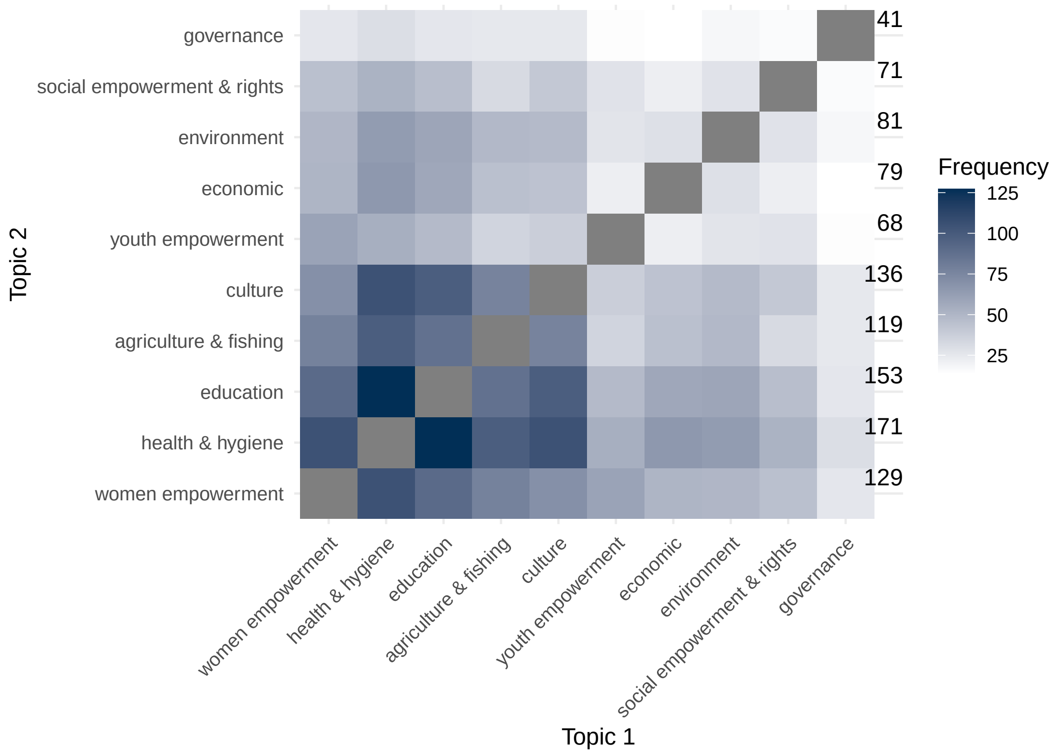


Figure I.3: Correlations in radio topics based on radio compendia

Note: The Figure shows the correlation between topics discussed by different radio stations.

(a) **Protagonists** (b) **Format** (c) **Audience**

Figure I.4: Word clouds of protagonists, format, and audience of radio Shows

Note: The wordcloud is created by pooling all words or sentences that were manually coded to be related to the respective topic. These are pre-processed by removing stop words, changed to lower case letters, and stemed. Then 1, 2, and 3 grams are created and used to plot the wordcloud based on term frequencies. Words are scaled by the square root of their frequency.