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

Felix Rusche
University of Mannheim

August 26, 2024

Job Market Paper (Draft)

Abstract

Can grassroots media serve as an effective policy tool for development? I investigate India's 2006 community radio policy which permits educational institutions and NGOs to obtain radio licenses with the purpose of promoting local development. For this, I collect unique data on the content and coverage areas of >250 radios. Through topic modeling and GPT-based content analyses, women's empowerment is identified as a key theme in radio programming. To causally investigate the effects of radio on women's empowerment, I leverage topography-driven variation in radio access in combination with a novel approach to reduce attenuation bias in randomly jittered survey coordinates. Exogenous exposure to community radio enhances girls' education, delays marriage, reduces fertility, and increases young women's autonomy.

Keywords: Mass Media, Policy, Women Empowerment, Education, Fertility, Radio
JEL Classifications: O12, J13, J18, L82, J16

Author: Felix Rusche, Department of Economics, University of Mannheim; felix.rusche@uni-mannheim.de
Acknowledgments: I benefited from discussions and feedback from Camille Urvoy, David Yanagizawa-Drott, Rachel Glennerster, Yoko Okuyama, Ruben Enikolopov, Doug Gollin, Philip Ager, Wladislaw Mill, David Atkin, Gianmarco León-Ciliotta, Elliot Motte, Christoph Rothe, Cristina Bellés-Obrero, Johannes Bauer, Patrick Sun, Johannes Gallé, Carsten Trenkler, Valentin Lang, Nikola Noske, 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, Royal Economic Society Annual Conference 2024, Federal Communications Commission (FCC) Economic Seminar 2024, 2024 Asia Meeting of the Econometric Society, Nordic Conference in Development Economics 2023, German Development Economics Conference 2023, German Economic Association Annual Conference 2023, Universitat Pompeou Fabra, South Asia Institute (Heidelberg University), Frankfurt-Mannheim Ph.D. Workshop, Barcelona School of Economics Ph.D. 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) and Joachim Herz Foundation is gratefully acknowledged. I also acknowledge support by the state of Baden-Württemberg through the 'bw High-Performance Computing Cluster' (bwUniCluster/bwHPC).

1 Introduction

Life opportunities of girls and women continue to be constrained by societal attitudes. In developing countries, the lack of opportunities is exemplified by women's relatively low educational attainment and high rates of early childbearing (Duflo, 2012). The media has been found to change attitudes and behaviour when listeners can relate to stories and characters (DellaVigna and La Ferrara, 2015). Whether the media can also be an effective policy tool to change the societal attitudes that constrain girls and women is less clear. Media content directly produced or distributed by the government may be perceived as slanted or sectarian. An alternative media policy has been implemented by the Indian government in 2006. The policy incentivizes local educational institutions and NGOs to establish community radio stations to broadcast locally produced content. Community radio stations are required to address development issues that are relevant for their local audiences. However, within this mandate, they make their own editorial choices. The programs of community radio stations have since strongly focused on education and women's empowerment – which are widely viewed as two of the country's main challenges (Inglehart et al., 2014).

To answer whether community media can be used as a policy instrument for development, I collect detailed information about the location, coverage area, launch date, and content of the >250 community radios launched by 2020. Starting with content, I transcribe and translate >5k radio shows and analyze their content using a Latent Dirichlet Allocation (LDA) topic model. I identify women's empowerment and education as key topics in radio programming. GPT based content analyses further show that radios discuss issues related to women's empowerment in a 'progressive' way. Specifically, these air content in favor of girls' education and family planning, and against child marriage and domestic violence. Next, I combine coverage areas with the National Family and Health Survey (NFHS), a large-scale demographic survey. To identify effects of radio, I exploit the topographic variation between radio towers and individuals in combination with a novel method to reduce attenuation bias in randomly jittered coordinates of the underlying survey data. More precisely, I improve on an important caveat of NFHS/DHS data. In such data, the reported coordinates of survey locations are randomly jittered prior to being made available to researchers. I draw on the publicly known jittering algorithm to compute the probability weight on the treatment area conditional on observing a jittered location. Moving to the results, I show that exogeneous exposure to CRS increases individuals' radio consumption and, importantly, consumption of development-related radio programs. Further, the results show that community radio is related to strong changes in variables indicative of women's empowerment. Girls and women exogenously exposed to CRS spend more time in school, delay their marriage, and lower their fertility. Young women also report higher autonomy with respect to decision-making and mobility.

CRS are an important part of the media environment in much of the developing world (Fraser and Estrada, 2001). India may be particularly suited to profit from community radio: around a quarter of the population remains illiterate, while 15% of men and 25% of women lack access to any mass media (IIPS and ICF, 2017). Moreover, the country's cultural and linguistic diversity make it well suited for CRS, which tailor their content to local audiences. Following a 1996 supreme court ruling and almost a decade of pressure from civil society, India passed a CRS policy in 2006. The legislation is restrictive with respect to who can obtain a license (NGOs and educational institutions) and requires the CRS to focus on local development while barring them from discussing news. By 2020, a total of 289 radios had been launched.

To get an idea of radios' content, I crawl, transcribe, and translate 5k+ radio shows from a website that allows radios to share content with one-another. Using Latent Dirichlet allocation (LDA), a topic model, I show that CRS discuss various topics related to development, the two largest of which are education and women's empowerment. To identify the effects of CRS on women's empowerment, I exploit topographic features between radio towers and receivers, a well-established identification strategy for radio and television signals.[1] However, the underlying NFHS/DHS data poses a major challenge when combining it with rather small treatment areas: to preserve the privacy of respondents, the survey coordinates are jittered by up to 5km prior to reporting them to researchers.[2] This is carried out by randomly drawing both direction and distance from two independent uniform distributions. The jittering introductes substantial measurement error and, hence, attenuation bias. I improve upon it by computing the expected coverage of each survey cluster conditional on the jittering algorithm and the observed coordinate.

Moving to results, I begin by asking whether individuals exposed to exogenous variation in access to community radio are more likely to listen to radio programs. The results show increases in both the consumption of radio and the propensity to have heard development-related messages on the radio. CRS thus reach an audience and, importantly, increase listeners' exposure to messages typically produced by such radios. I find no evidence for substitution away from other media, such as television, internet, or newspapers.

Moving to variables related to women's status, I start with educational outcomes. The results show that radios increase years of schooling, attendance rates, and the propensity to have obtained a degree. These effects are mainly driven by girls. The results suggest that living in an area exogenously exposed to radio over the entire time period makes young women around 3% more likely to have a primary, secondary, and higher degree respectively. The findings are driven by an increase in the value ascribed to girls' education

[1] For example, see Olken (2009); Yanagizawa-Drott (2014); Adena et al. (2015); Armand et al. (2020); Adena et al. (2020); Armand et al. (2020)

[2] 1% of rural observations by up to 10km

and a lower dropout rate due to early marriage. I do not find consistent effects of similar magnitude for boys. This implies that CRS increase both girls' status and parents' total investments in education.

CRS also have effects on the marriage market and fertility. Females exposed to radio are less likely to be married between the ages of 13-25. Effects for males are lagged, likely explained by the age gap between husbands and wives in India. Regarding fertility, I find decreases in the number of children women bear up to the age of around 35. 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.

Finally, I test whether CRS affect women's autonomy and variables on domestic violence. Young women exogenously exposed to radio are substantially more likely to have a say in household decisions and regarding their own mobility. Similarly, men also adjust their attitudes toward the autonomy of women, increasing the share of decisions in which they believe women should participate. In addition, I find suggestive evidence of decreases in women's attitudes toward domestic violence and the experience thereof. However, the effects are largely insignificant and, hence, rather suggestive. Results for male attitudes suggest no effect, meaning there is no evidence for a 'male backlash' to changes in women's status (Guarnieri and Rainer, 2021).

I rule out multiple potential threats to identification. First, results may be driven by arbitrary thresholds and choices regarding the regression's specification. To exclude this possibility, I vary the distance cut-offs for data inclusion and definitions of the treatment and outcome variables. The results remain when doing so. Furthermore, the results may be driven by omitted variable bias or pretrends correlated with the variation in radio exposure exploited in this paper. I test this in several ways. First, I follow Yanagizawa-Drott (2014) and regress variables related to women's status but unlikely to be affected by radio on exposure to radio. These include scheduled caste shares, travel times to the nearest city, urbanity, population density, and others. I find no effect of radio exposure on such variables, speaking against omitted variable bias. To test for pretrends, I further run regressions on education for age groups whose choices are unlikely to have been altered by radio. This includes cohorts that had already finished their educational choices when the first radios launched in 2005. I find no effects of radio exposure on these age cohorts, suggesting no pretrends in exposed areas. Finally, to test for both omitted variable bias and pretrends, I re-run all main regressions on a placebo sample. The placebo sample includes individuals in the vicinity of a radio station that launches after data collection, that is, after 2015. Individuals in the placebo sample will therefore receive a community radio station at a later point in time with no effects expected at the time of data collection. Placebo regressions show no effects on outcomes related to women's empowerment. Overall, the robustness and exogeneity checks suggest that neither pretrends nor omitted variable bias drive the results, stengthening the results' causal interpretation.

The article is most closely related to studies on the use of media as a policy for development and, in particular, with the purpose of empowering women. This encompasses the evaluation of government campaigns (Khalifa, 2022; de Walque, 2007) and of the effects of a radio station set up by the US occupying force in Japan on women's political participation (Okuyama, 2023). Most closely, this paper relates to an RCT by Glennerster et al. (2021) who test the effects of exposure to community radio in Burkina Faso against exposure to messages on contraception use produced by a British NGO. While they document positive effects of the latter, increasing exposure to community radio alone leads listeners to hold more conservative gender norms. In contrast, I document increases in variables typically associated with women's empowerment. These contrasting findings are likely driven by the fact that CRS in India are launched within a framework specifically aimed at fostering local development while being restrictive with respect to who can obtain a license. In Burkina Faso, community radios do not operate in such a framework, radios are run by a range of different organizations and not (typically) aimed at social development. Overall, my contribution and the key distinguishing feature of this article is the provision of evidence on the *intended large-scale and long-term* use of *media as a policy by and for a developing country.*

More broadly, this paper links with the literature on the effects of the media on socioeconomic outcomes (for reviews see La Ferrara, 2016; DellaVigna and La Ferrara, 2015; Enikolopov and Petrova, 2017). In the context of this article, the literature can broadly be categorized into two main branches: first, a number of papers study the unintended effects of entertainment media using observational data. For example, La Ferrara et al. (2012) show that the expansion of telenovelas throughout Brazil decreased fertility, mainly driven by exposing women to different ways of life. Other studies complement this evidence, showing that entertainment media can increase divorce rates (Chong and La Ferrara, 2009), increase schooling (Jensen and Oster, 2009), or reduce teen pregnancies (Kearney and Levine, 2015). A second strand of the literature test the effectiveness of exposing individuals to specific television programs in field experiments. For example, Banerjee et al. (2019) invite Nigerians to the screening of an MTV show featuring information on HIV/AIDS. Treatments like these can help to convey information and change attitudes and behaviors (Bernard et al., 2014; Berg and Zia, 2017; Arias, 2014; Ravallion et al., 2015; Green and Vasudevan, 2018; Cassidy et al., 2022; Kasteng et al., 2018; Murray et al., 2015), although some also find no or unintended negative effects (Coville et al., 2019; Bjorvatn et al., 2020). Both of these strands have in common that they are difficult to directly translate into larger policy frames: The first strand analyzes unintended consequences of entertainment media while the second strand mostly revolves around single movie or television programs. I contribute by evaluating the translation of the above evidence into a policy, namely the large-scale and intended use of grassroots or community radio for local development. In this context, I also broadly link to studies on community participation in development (Casey, 2018; Björkman and Svensson, 2009; Olken, 2007), providing evidence on community participation in local media.

Finally, I methodologically contribute to the literature using DHS, MISC, and other geocoded data with jittered locations. The DHS is likely the most used repeated survey in development economics and widely used in other fields, such as public health. Since 2005, Google Scholar lists >17k results for DHS and around 6.5% of all articles in both the Journal of Development Economics and World Development cite DHS surveys.[3] Furthermore, 117 articles in Economics' 'Top 5' journals cite the DHS since 2005. Previous research treated jittered coordinates as if they were correctly observed (e.g. Hjort and Poulsen, 2019; Guarnieri and Rainer, 2021). And despite measurement error being introduced through the jittering, this issue "[remains] largely unexplored in the development economics literature, despite the proliferation of research on accuracy and measurement error in household survey data", as Michler et al. (2022, p. 3) note. Their paper tests the effect of measurement error in studies on the relationship between weather and agricultural productivity. Given the low resolution of weather data, they find a negligible effect, but note that higher resolution data is likely "[...] to be sensitive to some spatial anonymization techniques" (p. 2). I propose a novel method that can substantially reduce measurement error when working with high resolution data and small treatment areas. The correction I propose substantially affects point estimates and levels of significance. On average, the point estimates increase by more than 50% when making the correction. Although standard errors slightly increase, the correction substantially affects levels of significance. These results are in line with a substantial decrease in attentuation bias due to measurement error in the treatment variable. The insights are relevant for any study using DHS or other data sets with jittered data. They are particularly relevant for studies with small treatment areas.

2 Context and Policy

2.1 Community Radio

Radio remains one of the most accessible media for people in developing countries. It is cheap, accessible to illiterate populations, low-tech, but can is also easily translated

[3]As of Feburary 12, 2024, >17k articles from 2005 mention *"demographic and health survey" OR "demographic and health surveys"*. Google Scholar lists 2,330 articles in the JDE published since 2005 (*source:"Journal of Development Economics"*). Of these, 150 mention *"demographic and health survey" OR "demographic and health surveys" OR "DHS"* and 158 either the DHS or MISC (OR condition extended by *"MISC"*). For World Development, a total of 6,610 articles are listed of whom 419 mention the DHS and 441 the DHS or MISC. Furthermore, 117 articles in Economics' 'Top 5' journals cite the DHS and 174 the DHS or MISC.

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). CRS 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 CRS, many countries, especially across Africa and Latin America, have granted licenses to a large number of CRS. For example, 93% of the villages in northern Benin had access to at least one CRS in 2009 (Keefer and Khemani, 2016). Boas and Hidalgo (2011) count 2,328 CRS in Brazil in 2008. In South Asia, where media is typically more strongly controlled by the state, CRS have only more recently started to gain pace (Raghunath, 2020).

India may be particularly suited to benefit from CRS. 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 the mass media, such as television, radio, cinema or newspapers (IIPS and ICF, 2017).

2.2 Community Radio in India: Policy

By the 1990s, India's state-run All India Radio (AIR) covered about 99% of the population. Regularly 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 allowing the setup of CRS in 2006 (Pavarala and Malik, 2007). Compared to other countries, the regulation of CRS 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 setup CRS: educational institutions, NGOs, and Krishi Vigyan Kendras (KvK). KvK are government-financed agricultural centers that aim 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. This means that local or state governments are usually not involved in deciding whether or not a radio is established. There are two key bottlenecks in the application process. 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 (information based on an expert interview with the MOIB and Raghunath (2020)).[4]

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 bans radios from airing (political) news.[5] Further, it holds radios responsible for spreading demeaning content about minorities and disadvantaged groups, such as women (Govt. of India, 2006).

To obtain funds, CRS can run 5 minutes of advertisements per 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).

3 Data and Descriptive Statistics

3.1 Data Collection and Preparation

Community Radio Stations Data on CRS is collected from a variety of sources. First, a list of all 289 CRS as of March 31, 2020 is obtained from the Ministry of Information and Broadcasting (MOIB).[6] Apart from the address and launch date, the list shows that 49% of CRS 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).

Stations are geolocated through rigorous web search using information on their name, address, and the license holder. The MOIB also provided me with a list including approximate locations (1.2km precision), which I used to verify the collected information.

[4]Specifically, the Community Radio Forum, a think tank of advocates, and the community radio association, an interest group of radio operators are part of the screening committee (Raghunath, 2020).

[5]Radios can, however, air newscasts produced by AIR though, according to multiple expert interviews, this only very rarely done in practice (Myers, 2011).

[6]The list can be accessed here (MOIB)

In total, 276 of 289 stations were verified as operational of which 92% or 264 stations were precisely geocoded (see G in Supplementary Material 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.[7,8]

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 CRS. Given that many clusters are out of reach of any radio signal, 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.[9] Additionally, different 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 CRS. The placebo sample, on the other hand, includes observations within 50km of a radio launched post-2015. Figure 1 shows the included and excluded data for the main sample.

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 2 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

[7]Radio coverage areas are calculated using the ITM algorithm through `cloud.rf`'s API.

[8]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.

[9]Figure B.3 in Appendix.

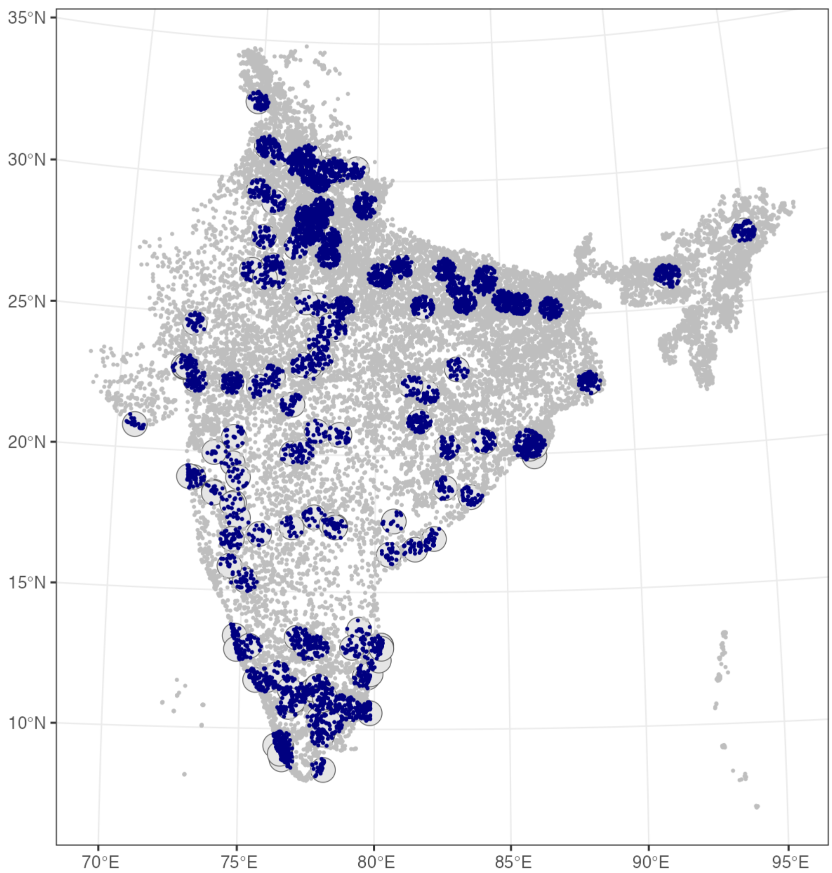


Figure 1: Visualization of NFHS data included (blue) and excluded (grey) in pre-2016 sample.

Observations within 50km of a given station are included, a total of 8,217 clusters. Each cluster includes around approximately 21 households. Colors indicate clusters' probability to be covered by a radio signal.

radio. For a developing country, these numbers are rather low. This is likely explained by the early and widespread introduction of television across 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. 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.

Next, I get idea for how men stand 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. These are only available for a small subsample of women and only collected in about 35% of the 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 agruing 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. Interestingly, 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 5. The first set of variables pertains 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).[10] Propagation controls further include the (expected) distance to the nearest radio tower (also see Chapter 4). In addition, I compute the travel time from each observation to

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

the nearest radio tower using Google's Direction API.[11] Finally, additional geographic controls cover the urbanization, population density, travel times to the nearest city[12], and distances to water bodies and national borders.

3.2 Descriptives: Content

Depending on the audience and aim of the institution running a particular radio station, CRS 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 CRS. 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 CRS (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 CRS. They provide a one-page fact sheet on each participating 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 radios' websites (if available). In total, I collect content information on 248 radios.[13] After identifying the main topics, I go through all the texts, manually marking words related to different topics (see Section **??** in 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 (see Section H in Appendix for more information).

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 programmes on 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. However, leadership roles continue to be dominated by men (SMART, 2023).

To get an idea of what radios are talking about, I crawl all >14k radio shows uploaded

[11] The data is visualized in Figure B.6 (in Appendix).

[12] 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).

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

to edaa.in, a platform where community radios can upload and exchange content. Using Google's Speech-to-Text API and Google Translate, I transcribe and upload 5,869 shows from 95 stations which uploaded content to the website.[14] 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).[15] The resulting topics are hand-labeled based on each topic's 15 most predictive terms (see Tables I.4 and I.5 in Supplemental Material). To get an idea of the content, I first collapse the topics into 8 categories.

The graph on the left of Figure 2 visualizes the average radio's share of development-related content across topics. 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.

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

[15]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.

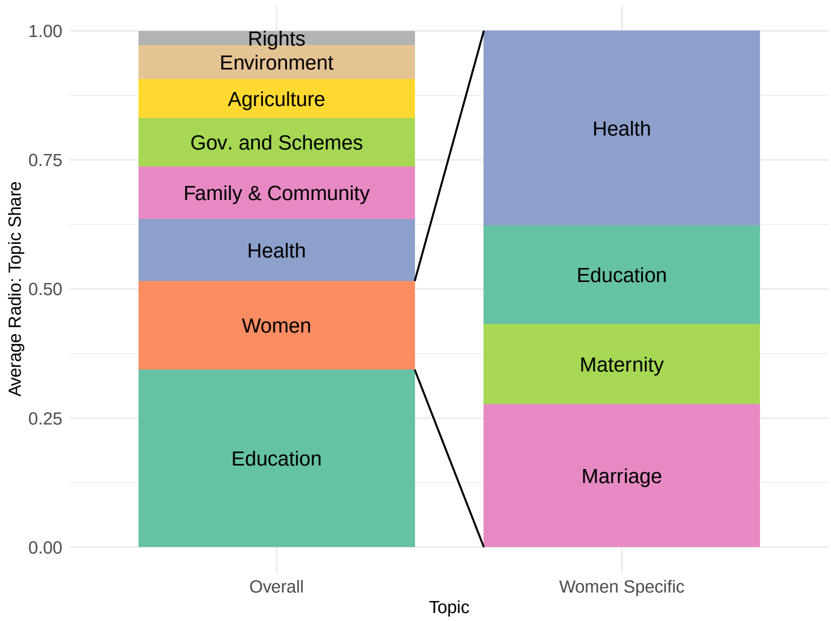


Figure 2: Radios' Share of Content Across Topics Based on LDA
Note: The above figure visualizes the distribution of topics of the average radio station. For this, translated transcripts of radio shows are assigned topic shares using an LDA model. Next, the average transcript is computed by station. Finally, 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 using a multistage evaluation of transcripts 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 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 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 is in favor, neutral or against the respective issue (e.g. child marriage or girls' education). Chapter I.2 in Supplementary Material 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 'progressive' in the sense that these argue infavor 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 'progressive' viewpoints 423 times.[16] Only two 'conservative' points of view are identified, with another 18 taking no or a neutral position. Overall, this suggests that the content produced by radios can be described as 'progressive'.

Overall, both sources on radios' content suffer from potential drawbacks. For instance, radios may upload selective shows or report selective topics they focus on. They may also shift their focus over time. However, it seems rather unlikely that within development related content, radios would

both topic analyses show that women-specific programs are a vital element of CRS' content. They further show that content produced on women's empowerment is progressive in the sense that radios produce content in favor of girls' education, later marriage, lower fertility, and against domestic violence.

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

Variable	cluster	n	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
CONTROLS: 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

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)
CONTROLS: ADD. GEO.		
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 2: Variable Descriptions and Sources

4 Spatial Jittering

A key component of the treatment variable of this study describes whether or not an individual lives in the treatment area, i.e., whether she is covered by a radio signal. As in many other developing countries, one of the most comprehensive sources of individual-level outcomes are DHS surveys.

In the DHS/NFHS surveys, enumerators gather precice coordinates of the central point of each enumeration area. These areas are small geographical units that typically cover about 20 households in India. However, to ensure privacy, the actual coordinates are not disclosed. Instead, the DHS adjusts the true locations by jittering them within a range of 0 to 2km in urban areas and 0 to 5km in rural areas.[17] This introduces substantial measurement error and as I will show below, some bias in studies drawing on geographic coordinates of DHS observations. This is particularly the case if treatment areas or distances to treatments are small relative to the displacement.

The jittering follows the "random direction, random distance" method (for a detailed description see Burgert et al., 2013)[18]:

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/rural) 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 administrative unit (e.g. district or state), a new location is drawn until the draw results in a location within the given administrative unit.

The PDF of drawing location x^* conditional on original location x can be characterized as follows:

[17] 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.

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

$$f(x^*|x) = \frac{I(A(x) = A(x^*)) \times I(d(x,x^*) \leq \bar{d})}{d(x,x^*)} / \int_{\hat{x}} \frac{I(A(x) = A(\hat{x})) \times I(d(x,\hat{x}) \leq \bar{d})}{d(x,\hat{x})} d\hat{x} \tag{1}$$

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

$$\text{where, } w(x) = \int_{\hat{x}} \frac{I(A(x) = A(\hat{x})) \times I(d(x,\hat{x}) \leq \bar{d})}{d(x,\hat{x})} d\hat{x} \tag{3}$$

$A(.)$ 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' 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 in combination with the observed jittered location x^* can be used to estimate with what probability a location observed at location x^* originally comes from the treatment area:

$$\Pi(T = 1|x^*) = \frac{\int_x f(x|x^*) \times T(x)dx}{\int_x f(x|x^*)dx} \tag{4}$$

where $T(x)\epsilon\{0, 1\}$ is the treatment status of location x. This hinges on $f(x|x^*)$, which can be derived using Bayes' theorem for random variables. More precisely:

$$f(x|x^*) \stackrel{(1)}{=} \frac{f(x^*|x)f(x)}{\int_x f(x^*|x)f(x)dx} \stackrel{(2)}{=} f(x^*|x) \tag{5}$$

Step (1) follows from Bayes' Theorem. Step (2) follows by assuming $f(x)$ to be uniform (Altay et al., 2022). In effect, this is equivalent to being agnostic about the DHS's survey sampling within small areas, i.e., making no assumptions about which direction or distance the x^* is more likely to have come from. This is consistent with the random distance, random direction jittering algorithm and essentially translates to not making further assumptions about which small geographic areas enumerators are more likely to have been to. Importantly, this step does not assume that the DHS/NFHS is equally likely to choose areas to survey across India. Rather, information on the DHS's choices are preserved in the observed locations x^*. For instance, if the DHS conducted a random draw of individuals across India, the distribution of x^* would follow the population distribution, given the "random direction, random distance" jittering method.[19]

$\Pi(T = 1|x^*)$ can then be written as follows:

$$\Pi(T = 1|x^*) = \frac{\int_x \frac{T(x)}{d(x,x^*) \times w(x)} dx}{\int_x \frac{1}{d(x,x^*) \times w(x)} dx} \tag{6}$$

and computationally be implemented as follows:

For each location $x*$:

1. create an equidistant grid of points x within distance $\bar{d}$ and administrative unit $A(x^*)$
2. for each x: generate a second grid with points z at distance $\bar{d}$ of x and in adminstrative unit $A(x^*)$ and compute
   $w(x) = \sum_z \frac{1}{d(x,z)}$.
3. estimate: $\Pi(T = 1|x^*) = \frac{\sum_x T(x)/(w(x) \times d(x,x^*))}{\sum_x 1/(w(x) \times d(x,x^*))}$

The result $\Pi(T = 1|x^*)$ 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 the idea. The left hand figure shows the entire PDF and the right hand side the probability mass on original locations within the treatment area.

[19] While the most agnostic approach with respect to DHS's decisions in the field, a potential improvement to this would be weighting $f(x)$ using high-resolution population data and/or by estimating which areas the DHS defines as urban or rural. However, given that both are subject to measurement error and author assumptions regarding the DHS's decisions, this paper's author prefers the more agnostic approach. For example, weighting $f(x)$ using high-resolution population data would assume that the DHS/NFHS's sampling frame follows the population distribution. While this might be a reasonable assumption, the DHS oversamples some populations, as indicated by sample weights. Therefore, following the population-weighted approach would require additional information or assumptions about the survey weights and its interaction with high resolution population data.

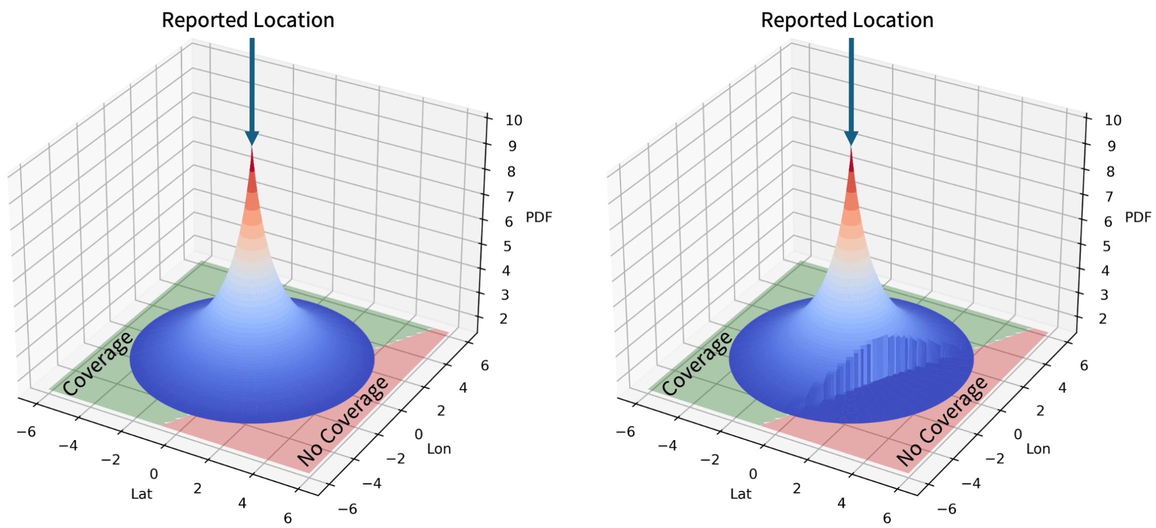


Figure 3: Visualization of Jittering Correction

Note: The figures above show the PDF of x conditional on observing x^* (Reported Location). 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.

4.1 Expected (Squared) Distances

The jittering does not solely affect the measurement of treatment status. Importantly, it also affects the calculation of distances between observed locations x^* and other points of interest. In this study, this is particularly relevant for the control of the distance between an observed location x^* and a given radio tower t. In other studies, this is relevant to compute distances to treatment areas (such as in RDD settings), schools, Christian missions, public services, etc.

First note that when computing the distance between an observed location x^* and a radio tower t, the observed distance based on x^* and t generally does not equal the expected distance when taking into account the jittering algorithm. To see this, first note that:

$$\mathbb{E}(d(x,t)) = \int_x f(x|x^*)d(x,t)dx \tag{7}$$

The computation can be performed using the box provided above by simply computing a distance between x and t in Step 2 and replacing $T(x)$ by $d(x,t)$ in Step 3. For locations whose jittering is unaffected by an administrative border, one can further derive a closed-form solution for both the expected and expected squared distance.

This can be done using donuts. To see this, first note that displacement is uniform in direction and distance. Thus, if one were to split the 5km circle around the reported loca-

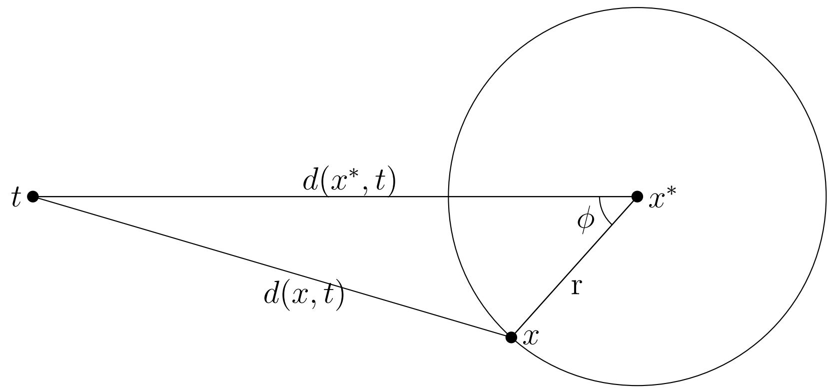


Figure 4: Visualization of setup and Law of Cosines

tion 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 original location x that lies on this donut as shown in Figure 4. To calculate the distance $d(x,t)$ between x and a radio tower t, one can draw on the Law of Cosines, which states $d(x,t) = \sqrt{r^2 + d(x^*,t)^2 - 2rd(x^*,t)cos(\phi)}$. 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 x on the circle. Intuitively, this moves x in infinitesimal steps once around the circle. At each step, $d(x,t)$ is calculated and, overall, expectation:

$$\mathbb{E}\left(d(x,t)|r\right) = \frac{1}{2\pi}\int_0^{2\pi} \sqrt{r^2 + d(x^*,t)^2 - 2rd(x^*,t)cos(\phi)}\, d\phi \tag{8}$$

$$= \frac{2}{\pi}(r + d(x^*,t))^2 E\left(\frac{2\sqrt{rd(x^*,t)}}{r + d(x^*,t)}\right) \tag{9}$$

Equation 9 follows by rewriting the equation as a function of the elliptic integral of the second kind $E(\cdot)$, 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).

Given that r is generally unknown and anywhere between 0 and $\bar{d}$ the expected distance between x and t can be obtained by integrating over all donuts within distance $\bar{d}$ with uniform priors, as suggested by the "random distance, random direction" jittering algorithm:

$$\mathbb{E}(d(x,t)) = \frac{1}{\bar{d}}\int_0^{\bar{d}} \frac{2}{\pi}(r + d(x^*,t))^2 E\left(\frac{2\sqrt{rd(x^*,t)}}{r + d(x^*,t)}\right) dr \tag{10}$$

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.1, 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^* exactly equals 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. Second, the absolute and relative difference between $d(x^*, t)$ and $\mathbb{E}(d(x, t))$ decreases in $d(x^*, t)$. However, $d(x^*, t)$ is always smaller than $\mathbb{E}(d(x, t))$. Third, given that the difference increases in $\bar{d}$, differences for urban areas are smaller.

Figure A.2 further compares the theoretical result in Equation 10 to results obtained when empirically estimating expected distances using the grid in the above box. As expected, for observations whose jittering is unaffected by an administrative border, the empirical results equal those of the theoretical result. For those affected by a border, the distances are scattered around the theoretical line.

Finally, the expected squared distance can be derived following the same logic as above:

$$\mathbb{E}(d(x,t)^2) = \int_0^{\bar{d}} \frac{1}{2\pi} \int_0^{2\pi} r^2 + d(x^*,t)^2 - 2rd(x^*,t)cos(\phi)\, d\phi dr \tag{11}$$

$$= d(x^*,t)^2 + \frac{1}{\bar{d}} \int_0^{\bar{d}} r^2\, dr \tag{12}$$

$$= \begin{cases} d(x^*,t)^2 + 8.\bar{3} \text{ in urban clusters} \\ d(x^*,t)^2 + 1.\bar{3} \text{ in rural clusters} \end{cases} \tag{13}$$

The above equation shows that 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. In relative terms, this again means that the difference is much higher for low $d(x^*, t)$ as visualized in Figure **??**.

Overall, the results regarding expected (squared) differences 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.

5 Empirical Strategy

To identify the causal effect of community radios, variation in coverage due to local topographical features is exploited (Olken, 2009).[20] This is done in several steps: First, using the irregular terrain model (ITM, 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 CRS is estimated.[21] 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 (Yanagizawa-Drott, 2014).[22] The remaining variation in the signal strength 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. 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. 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 surroundings 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 CRS, 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:

$$Exposure_{c(i)} = \sum_{r=1}^{R} AddedCoverageProbability_{c(i),R} \times f(Timeshare_R) \qquad (14)$$

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

[21]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.

[22]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.

where $f(Timeshare_R)$ is a function of the share of time between 2005 and 2015 that radio R was on air where $Timeshare_R$ ranges from 0 to 1. $AddedCoverageProbability_{c(i),R}$ describes the increase in probability to be covered by a radio signal that radio R brings (ranging from 0 to 1) in addition to previously launched radios. Therefore, if individual i is covered with full probability by a radio that launches in 2005 ($AddedCoverageProbability_{c(i),1} = 1$) and a second radio launches in t=2 and covers the location with some nonzero probability, then $AddedCoverageProbability_{c(i),2} = 0$. As a result, $(\sum_{r=1}^{R} AddedCoverageProbability_{c(i),R}) \epsilon [0, 1]$ and $Exposure_c(i) \epsilon [0, 1]$. For example, a value of $Exposure_c(i) = 1$ means that the radio covers individual i with full probability and is on air from 2005 to 2015.

This assumes a linear treatment effect over time. This may be the most agnostic approach given that it is hard to form expectations on the 'true' functional form, which may further differ from outcome to outcome. I explore alternative functional forms in Chapter C in Supplementary Material. The results suggest that the effect of radio may be nicely resembled by a quadratic effect over time. Though I keep the linear effect as the default, I repeat all regressions using a quadratic effect over time, i.e., replacing $Timeshare_R$ by $Timeshare_R^2$ in Equation 14.

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

$$y_i = \beta Exposure_{c(i)} + Distance_{c(i)}\delta + Geography_{c(i)}\omega + X_i\lambda + \gamma_{r(i)} + \epsilon_{i,c(i),r(i)} \tag{15}$$

here, y_i is the outcome of interest for individual i. $Distance_{c(i)}$ includes second-order distance polynomials to the closest, second, and third closest radio towers for i's cluster $c(i)$.[23] Further, I control for the travel time between the cluster and the closest radio tower.[24] $Geography_{c(i)}$ includes topography controls, i.e. second-order polynomials of ruggedness and altitude. In addition, X_i includes several individual-level covariates and variables related to clusters' surroundings, such as population density or travel time to the nearest city. 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 effect of treatment of interest, which describes the exposure of cluster $c(i)$ to CRS.

Identification relies on exogenous variation in exposure to CRS 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 γ_c effectively controls for any specific characteristics of the coverage area. This includes the fact that certain areas receive a CRS 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

[23] values for the second and third closest towers are capped at 50km.
[24] see Figure B.6 for a visualization

the transmitter and the receiver and uncorrelated with all other determinants of women's empowerment.

In addition to the variables mentioned above, X_i includes a number of controls: on the cluster level these include the log. population density, log. travel time to the nearest urban area[25] , proximity to national borders and water bodies, and whether the cluster is defined as urban by the NFHS.[26] On the individual level, I control for 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.[27]

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 Supplemental Material ??). This follows Durante et al. (2019), DellaVigna et al. (2014), Adena et al. (2020), and Olken (2009).[28]

6 Results

6.1 Exogeneity Check

I start by testing the treatment's correlation with pre-determined cluster characteristics that are unlikely to be affected by the radio. Here, I follow Yanagizawa-Drott (2014) and regress such characteristics on the treatment variable. The regressions differ from Equation 15 only includes distance and geography controls as well as CRS fixed effects. The reason is that the other controls serve as outcomes in the exogeneity regression.

The regressions reported in Table D.13 include outcomes potentially correlated with the role of women but unlikely to be affected by radio: population density, caste (SC/ST), Muslim, urbanity, proximity to national borders, travel time to nearest city, and travel time to the nearest radio tower. The regression on travel time to the nearest radio naturally excludes this variable from the set of controls.

The results are insignificant across all variables. This holds across different specifications, i.e. when assuming either linear and quadratic effects of radio over time. Overall, this strengthens the causal interpretation of the variation used in this paper. Neverthe-

[25] Travel times are based on Weiss et al. (2018)'s definition of high-density urban areas. They define these "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).

[26] The NFHS follows the 2011 Indian population census' definition of urban/rural, see Census of India (2011).

[27] Note that in regressions that only include women, variables on gender are irrelevant.

[28] Key packages used: Regressions: *fixest* (Bergé, 2018); Spatial operations: *sf* (Pebesma, 2018); Table Export: *modelsummary* (Arel-Bundock et al., 2023) (all in R)

less, it is important to note that all the outcomes used in the exogeneity regressions are included as controls in regressions below.

6.2 Radio Consumption and Content Reception

Next, I test 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 3.3pp. in the linear and 5pp. in the quadratic model. This corresponds to an increase of 17 to 27% percent compared to the baseline. Given that the baseline includes treated units, the actual relative effect is likely greater.[29]

Columns (3) and (4) provide evidence on development-related content. These variables get closest to measuring exposure to content typically produced by CRS, as suggested by the content analyzes above. 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 5.3 to 7.6pp. depending on variable and model. This suggests strong increases by 27 to 46% compared to baseline when fully exposed over the entire period of time. Regarding gender differences in radio and content exposure, Table D.1 (in the appendix) shows that the observed effects are greater in magnitude for women.

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 CRS, which attempts to bring communities or specific groups, such as women, together.

Finally, Table D.12 (in the appendix) tests the effects of CRS 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 and suggests that people do not stop consuming other media to listen to radio.

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

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

Panel A: Linear Effects Over Time

	Radio Owner	Radio Consumer	Radio Familyplanning	Radio HIV/AIDS
exposure	−0.010	0.033*	0.053**	0.065**
	(0.018)	(0.019)	(0.025)	(0.026)
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: Quadratic Effects Over Time

	Radio Owner	Radio Consumer	Radio Familyplanning	Radio HIV/AIDS
exposure2	−0.004	0.050**	0.062**	0.076***
	(0.016)	(0.020)	(0.027)	(0.029)
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

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 5. 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 5. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

6.3 Education, Marriage, and Fertility

Moving to the key variables related to evaluating effects on women's empowerment, the effects on three interrelated variables are investigated: education, marriage, and fertility. Early marriage has been shown to constrain women's education (in India) (Maertens, 2013). Education, on the other hand, has generally been linked to reduced fertility (Basu, 2002) while the lack thereof remains an important barrier to women's empowerment in India (Jensen and Oster, 2009).

6.3.1 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.

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)
is female = 1 x exposure	0.070	0.245***	0.392**	0.282	0.493**	0.309**
	(0.064)	(0.094)	(0.164)	(0.190)	(0.219)	(0.122)
is female = 0 x exposure	0.051	0.223***	0.139	0.121	0.195	0.178
	(0.059)	(0.082)	(0.116)	(0.187)	(0.197)	(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

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)
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

Table 4: Effects of CRS 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 5. 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 CRS. 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 during the entire period of time increases the probability for girls of obtaining a degree by 3pp at all levels of schooling. 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.

But why does education increase? To answer this, 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) show the results on reasons for dropout. Panel (A) starts with a simple regression of having dropped out of school on exposure. As expected, the results closely mirror those of school attendance, showing a fall in dropout rates for girls in higher and lower secondary education. Next, 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 in Panels (B) to (G) 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 (Panel B). Costs are also substantially less likely to be cited as a reason for dropout for girls in higher secondary school. The third main reason is that 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, often 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.

6.3.2 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 place rather than the age of marriage. The result confirms the above results on marriage

Panel A: Linear Effects Over Time

	Primary	Secondary	Higher
is female = 1 x exposure	0.032**	0.037*	0.027*
	(0.015)	(0.021)	(0.015)
is female = 0 x exposure	0.013	0.017	0.015
	(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 B: Quadratic Effects Over Time

	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

Table 5: Effects of CRS 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 CRS and given that they have been able to finish the degree. Panel A assumes linear effects over time and Panel B quadratic effects. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter 5. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

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 40% decrease in the average of the dependent variable when exposed over the entire time period. At 14%, the relative effect is lower but remains high for women between the ages of 19 and 24. For men, the results are similar 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. 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 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 exogeneous exposure to CRS results in substantial delays in marriage, including early marriage of girls.

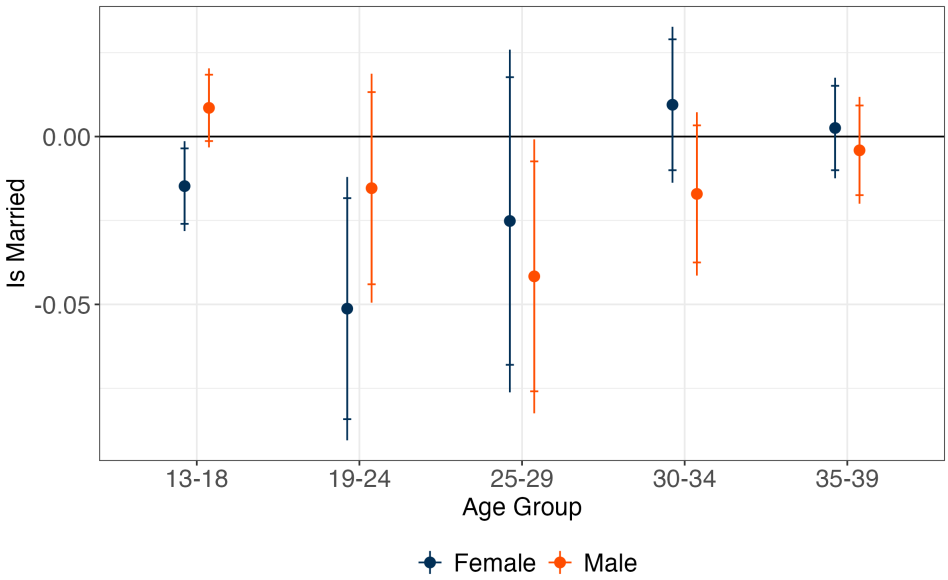


Figure 5: Exposure and Radio Consumption

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 5. Full regression results are shown in Table D.5. Unless otherwise specified, regressions include all applicable controls mentioned in Chapter 5. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010).

6.3.3 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 throughout the time frame 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.

6.4 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.

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	−0.001	−0.079**	−0.138**	−0.210**	−0.033	−0.012
	(0.003)	(0.034)	(0.069)	(0.098)	(0.073)	(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

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	−0.001	−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

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 5. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

Surveys on these are only conducted in around a third of survey clusters, meaning that the treatment variation available for identification is substantially reduced.

Panels (A) and (B) of Table 7 present regressions 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. The results show overall positive effects driven by young women between the ages of 15 and 25.[30] Panel (C) and (D) show a shift in men's views on their wifes involvement in household decisions as well. Other than for women, the coefficients are positive in 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.

Regarding attitudes, Table D.8 (in the appendix) shows regressions on whether women find it justifiable for their husbands to beat their wives under any circumstance in Columns (1) and (2). Columns (3) and (4) provide evidence on the number of reasons women list that justify domestic violence. The results are congruent with those on autonomy in the sense that coefficients suggest decreases in approval of domestic violence, especially among younger cohorts. However, they are also insignificant. The results for men suggest no change in attitudes (Table D.9 in the appendix).

Finally, Table D.10 tests the effects of exposure to any sexual, physical, or emotional violence from partner. Similarly to 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 sug-

[30]Tables D.6 and D.7 show separate results for women's autonomy with respect to decisions and mobility, respectively. These show that the effect is driven by both mobility and decision-making power.

gestive.

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 lack of a 'male backlash' may be explained by the nature of community radio which transmit information and views by community members. 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 CRS. 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).

6.5 Robustness and Placebo

In this section, I discuss robustness and placebo checks.

Starting with robustness 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 J (in the appendix) applies different standard errors, this time clustered at the subdistrict level. Further, Section K (in the appendix) varies the threshold of data inclusion by reducing it to observations within 40km of the radio tower (instead of 50km). The results are robust to any of the above changes.

More importantly, I test regressions for omitted variable bias and pre-trends in addition to the exogeneity check in Section 6.1. Starting with education, I can directly test for pre-trends by repeating regressions on school degrees and years of education for individuals that have likely finished their educational choices by the time the first radios arrive. Specifically, I repeat regressions for individuals above the age of 30 who were above the age of 20 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.

To more generally test for pre-trends and omitted variable bias in areas that obtain a radio signal, 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, starting 2016. I compute exposure, now defined between 2016 and 2020. 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). 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

Panel A: Women - Linear Effects Over Time

	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

Panel B: Women - Quadratic Effects Over Time

	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

Panel C: Men - Linear Effects Over Time

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-54)
exposure	0.048	0.058*	0.040	0.083*	−0.027
	(0.031)	(0.030)	(0.039)	(0.042)	(0.041)
Num.Obs.	30 580	10 766	8571	6862	4381
R2 Adj.	0.079	0.078	0.086	0.079	0.087
Mean Y	0.816	0.816	0.816	0.818	0.812
SD Y	0.285	0.284	0.282	0.287	0.292

Panel D: Men - Quadratic Effects Over Time

	Autonomy	Autonomy (15-25)	Autonomy (26-35)	Autonomy (36-45)	Autonomy (45-54)
exposure2	0.044	0.063*	0.032	0.078**	−0.038
	(0.033)	(0.034)	(0.043)	(0.038)	(0.043)
Num.Obs.	30 580	10 766	8571	6862	4381
R2 Adj.	0.079	0.078	0.085	0.079	0.088
Mean Y	0.816	0.816	0.816	0.818	0.812
SD Y	0.285	0.284	0.282	0.287	0.292

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 5. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

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 regression results for the placebo sample allowing for linear or quadratic effects over time. The regressions show no effects on outcomes. This speaks against pre-trends in areas to be covered by a community radio. It also speaks against omitted variable bias as a result of a correlation between the variation in radio exposure exploited in this paper and the outcome variables.

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

6.6 Evaluating the Jittering Correction

Finally, I compare the results in the main regressions presented in Sections 6.3 and 6.4 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.2 (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. A simple comparison is the number of significant estimates from the main paper that remain without the correction. Of the 16 significant coefficients, only seven remain. Further, only two of the 16 coefficients remain at the same level of significance. These numbers are driven by the size of coefficients rather than changes in standard errors. While the latter generally slightly increase when applying the proposed corrections, coefficients simultaneously increase substantially more. On average, their size grows 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.

The number of coefficients significant and at the same level of significance increase to eleven and three respectively when correcting for the expected distance instead of the simple line of sight (see Table E.3 in the appendix). However, coefficients remain around 53% smaller in size when compared to those in the main results.

While the simple comparison made above is clearly imperfect, it nevertheless suggests that the correction I propose substantially improves upon the attentuation bias introduced by the jittering of survey coordinates. 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.

7 Summary and Concluding Remarks

This paper evaluates India's 2006 community radio policy, which was established to further local development. Based on information gathered on the content of CRS, I focus on women's empowerment, one of the radios' main themes. For identification, I exploit topographic features in combination with a novel approach to reduce measurement error and bias in randomly jittered survey coordinates. The results show that CRS have substantial effects on the attitudes and behaviors of and toward women and girls. Areas exogeneously exposed to CRS show increased education and degree completion rates for girls. Young women marry later and have fewer kids. I also find evidence for greater autonomy of young women. Suggestive results further point toward changes in women's attitudes toward domestic violence, although the results are insignificant. I find no evidence indicating a male backlash to these changes.

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). In contrast to Glennerster et al. (2021), I find that community radio changes variables in a direction associated with women's empowerment. A key difference is that Burkina Faso's community radio operators are not subject to a policy framework aimed at development. This suggests that the Indian government's regulatory framework, under which licenses are distributed and content is produced, plays an important role in explaining why radios have the observed effect.

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 CRS have already taken first steps, e.g., by joining social media or broadcasting online.[31] 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 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 (Bureau, August 2024). It would therefore be important to expand the evidence for 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 an investigation of the channels driving effects of community radio. Specifically, the

[31]For example, `radio.garden` features many CRS around the world, including India.

effects observed may be driven by changes in information, attitudes, or beliefs about others' points of view (i.e. peer effects). Community radio may be particularly able to activate the latter. To investigate this, field work, such as through RCTs, may be a viable path. Finally, and in addition to the topical contribution, my paper also suggests a novel approach to deal with spatially jittered survey data. As I demonstrate, the correction strongly improves 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. 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.7

References

Adena, M., R. Enikolopov, M. Petrova, V. Santarosa, and E. Zhuravskaya (2015): "Radio and the Rise of The Nazis in Prewar Germany," *The Quarterly Journal of Economics*, 130(4), 1885–1939.

Adena, M., R. Enikolopov, M. Petrova, and H.-J. Voth (2020): "Bombs, Broadcasts and Resistance: Allied Intervention and Domestic Opposition to the Nazi Regime During World War II," *SSRN Electronic Journal*.

Altay, U., J. Paige, A. Riebler, and G.-A. Fuglstad (2022): "Fast Geostatistical Inference under Positional Uncertainty: Analysing DHS Household Survey Data," .

Anderson, J. and A. Lightfoot (2019): "The School Education System in India: An Overview," Tech. rep., British Council, New Delhi.

Arel-Bundock, V., J. Gassen, N. Eastwood, N. Huntington-Klein, M. Schwarz, B. Elbers (0000-0001-5392-3448), G. McDermott, and L. Wallrich (2023): "Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready," .

Arias, E. (2014): "Media, Common Knowledge, and Violence Against Women: A Field Experiment on Norms Change in Mexico," Working paper.

Armand, A., P. Atwell, and J. F. Gomes (2020): "The Reach of Radio: Ending Civil Conflict through Rebel Demobilization," *American Economic Review*, 110, 1395–1429.

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

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

Benoit, K., K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller, and A. Matsuo (2018): "Quanteda: An R Package for the Quantitative Analysis of Textual Data," *Journal of Open Source Software*, 3, 774.

Berg, G. and B. Zia (2017): "Harnessing Emotional Connections to Improve Financial Decisions: Evaluating the Impact of Financial Education in Mainstream Media," *Journal of the European Economic Association*, 15, 1025–1055.

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

Bernard, T., S. Dercon, K. Orkin, and A. S. Taffesse (2014): "The Future in Mind: Aspirations and Forward-Looking Behaviour in Rural Ethiopia," *CEPR Discussion Paper.*

Björkman, M. and J. Svensson (2009): "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda," *Quarterly Journal of Economics*, 124, 735–769.

Bjorvatn, K., A. W. Cappelen, L. H. Sekei, E. Ø. Sørensen, and B. Tungodden (2020): "Teaching Through Television: Experimental Evidence on Entrepreneurship Education in Tanzania," *Management Science*, 66, 2308–2325.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003): "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993–1022.

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

Bureau, U. C. (August 2024): "U.S. and World Population Clock," Tech. rep., U.S. Census Bureau.

Burgert, C. R., J. Colston, R. Thea, and B. Zachary (2013): "Geographic Displacement Procedure and Georeferenced Data Release: Policy for the Demographic and Health Surveys," Tech. Rep. 7, ICF International, United States Agency for International Development (USAID), Calverton, Maryland.

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

Casey, K. (2018): "Radical Decentralization: Does Community-Driven Development Work?" *Annual Review of Economics*, 10, 139–163.

Cassidy, R., A. Dam, W. Janssens, U. Kiani, and K. Morsink (2022): "Father of the Bride, or Steel Magnolias? Targeting Men, Women or Both to Reduce Child Marriage," Working Paper 22/50, Institute for Fiscal Studies.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2024): "On Binscatter," *American Economic Review.*

Census of India (2002): "Census of India : General Note," https://web.archive.org/web/20220314142033/https://www.censusindia.gov.in/Census_Data_2001/C

——— (2011): "Rural Urban Distribution of Population (Provisional Population Totals)," Tech. rep., Ministry of Home Affairs, New Delhi.

Chong, A. and E. La Ferrara (2009): "Television and Divorce: Evidence from Brazilian Novelas," *Journal of the European Economic Association*, 7, 458–468.

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

——— (2010): "Spatial Econometrics," in *Microeconometrics*, ed. by S. N. Durlauf and L. E. Blume, London: Palgrave Macmillan UK, The New Palgrave Economics Collection, 303–313.

Coville, A., V. Di Maro, F. A. Dunsch, and S. Zottel (2019): "The Nollywood Nudge: An Entertaining Approach to Saving," Policy Research Working Paper 8920, The World Bank, Washington D.C.

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

——— (2022): "About DAVP," http://crfc.in/about-davp/.

de Walque, D. (2007): "How Does the Impact of an HIV/AIDS Information Campaign Vary with Educational Attainment? Evidence from Rural Uganda," *Journal of Development Economics*, 84, 686–714.

DellaVigna, S., R. Enikolopov, V. Mironova, M. Petrova, and E. Zhuravskaya (2014): "Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia," *American Economic Journal: Applied Economics*, 6, 103–132.

DellaVigna, S. and E. La Ferrara (2015): "Economic and Social Impacts of the Media," in *Handbook of Media Economics*, ed. by S. P. Anderson, J. Waldfogel, and D. Strömberg, North-Holland, vol. 1 of *Handbook of Media Economics*, 723–768.

Duflo, E. (2012): "Women Empowerment and Economic Development," *Journal of Economic Literature*, 50, 1051–1079.

Durante, R., P. Pinotti, and A. Tesei (2019): "The Political Legacy of Entertainment TV," *American Economic Review*, 109, 2497–2530.

Enikolopov, R. and M. Petrova (2017): "Mass Media and Its Influence on Behaviour," Els Opuscles Del CREI 44, Centre de Recerca en Economia Internacional, Barcelona, Spain.

Enikolopov, R., M. Petrova, and E. Zhuravskaya (2011): "Media and Political Persuasion: Evidence from Russia," *The American Economic Review*, 101, 3253–3285.

Fraser, C. and S. R. Estrada (2001): "Community Radio Handbook," Handbook, UNESCO, Paris.

Fraser, C. and S. Restrepo-Estrada (2002): "Community Radio for Change and Development," *Development*, 45, 69–73.

Gius, E., J. C. Meister, M. Meister, M. Petris, M. Schumacher, and D. Gerstorfer (2023): "CATMA," Zenodo.

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

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

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

Grün, B. and K. Hornik (2011): "Topicmodels: An R Package for Fitting Topic Models," *Journal of Statistical Software*, 40, 1–30.

Guarnieri, E. and H. Rainer (2021): "Colonialism and Female Empowerment: A Two-Sided Legacy," *Journal of Development Economics*, 151, 102666.

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

Hjort, J. and J. Poulsen (2019): "The Arrival of Fast Internet and Skilled Job Creation in Africa," *American Economic Review*, 109, 1032–79.

Hufford, G. A. (2002): "The ITS Irregular Terrain Model, Version 1.2. 2, the Algorithm," Tech. rep.

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

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

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

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

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

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

Kasampalis, S., P. I. Lazaridis, Z. D. Zaharis, A. Bizopoulos, S. Zettas, and J. Cosmas (2013): "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)*, 1–6.

Kasteng, F., J. Murray, S. Cousens, S. Sarrassat, J. Steel, N. Meda, M. Ouedraogo, R. Head, and J. Borghi (2018): "Cost-Effectiveness and Economies of Scale of a Mass Radio Campaign to Promote Household Life-Saving Practices in Burkina Faso," *BMJ Global Health*, 3, e000809.

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

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

Khalifa, S. (2022): "Female Genital Cutting and Bride Price," Working Paper.

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

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

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

La Ferrara, E., A. Chong, and S. Duryea (2012): "Soap Operas and Fertility: Evidence from Brazil," *American Economic Journal: Applied Economics*, 4, 1–31.

Le Pennec, E. and K. Slowikowski (2023): "Ggwordcloud: A Word Cloud Geom for 'Ggplot2'," .

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

Michler, J. D., A. Josephson, T. Kilic, and S. Murray (2022): "Privacy Protection, Measurement Error, and the Integration of Remote Sensing and Socioeconomic Survey Data," *Journal of Development Economics*, 158, 102927.

MOIB (2020): "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.

Murray, J., P. Remes, R. Ilboudo, M. Belem, S. Salouka, W. Snell, C. Wood, M. Lavoie, L. Deboise, and R. Head (2015): "The Saturation+ Approach to Behavior Change: Case Study of a Child Survival Radio Campaign in Burkina Faso," *Global Health: Science and Practice*, 3, 544–556.

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

Nirmala, Y. (2015): "The Role of Community Radio in Empowering Women in India," *Media Asia*, 42, 41–46.

Okuyama, Y. (2023): "Empowering Women Through Radio: Evidence from Occupied Japan," .

Olken, B. A. (2007): "Monitoring Corruption: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy*, 115, 200–249.

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

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

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

Raghunath, P. (2020): "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, 145–192.

Ravallion, M., D. van de Walle, P. Dutta, and R. Murgai (2015): "Empowering Poor People through Public Information? Lessons from a Movie in Rural India," *Journal of Public Economics*, 132, 13–22.

Rusche, F. (2024): "Chapter 17: Large Language Model APIs," in *APIs for Social Scientists: A Collaborative Review*, ed. by P. C. Bauer, C. Landesvatter, and L. Behrens.

SMART (2023): "Gender Analyses in Community Radio Stations, 2023," Tech. rep., SMART NGO, New Delhi.

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

UNESCO (2013): "Statistics on Radio," .

Varshney, D., P. K. Joshi, A. Kumar, A. K. Mishra, and S. Kumar Dubey (2022): "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*, 160, 106067.

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

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

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

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

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

Online Appendix

Table of Contents

A Spatial Jittering: Expected Distance **47**

B Descriptives **50**

C Functional Form: Explorative Analysis **53**

D Additional Results **55**

E Results without Jittering Correction **68**

F Placebo and Robustness **71**

G Data on Radios: Data Gathering and Preparation **74**

H Radio Content I: Information from Compendia **74**

I Radio Content II: Information from Radio Show Recordings **77**

I.1 Data Preparation and Topic Model 77

I.2 GPT-Based Content Analysis 82

J Results Using Clustered Standard Errors **86**

K Alternative Specifications **89**

A Spatial Jittering: Expected Distance

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

$$E(x) = \int_0^{\frac{\pi}{2}} \sqrt{1 - x^2 \sin^2(\phi)} d\phi \tag{16}$$

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

$$\begin{aligned}
\mathbb{E}(d(x,t)|r) &= \frac{1}{2\pi} \int_0^{2\pi} \sqrt{r^2 + d(x^*,t)^2 - 2rd(x^*,t)\cos(\phi)} d\phi \qquad | \quad \text{By symmetry of circle} \\
&= \frac{1}{2\pi} \int_0^{2\pi} \sqrt{r^2 + d(x^*,t)^2 + 2rd(x^*,t)\cos(\phi)} d\phi \qquad | \quad \text{By symmetry of circle} \\
&= \frac{1}{\pi} \int_0^{\pi} \sqrt{r^2 + d(x^*,t)^2 + 2rd(x^*,t)\cos(\phi)} d\phi \qquad | \quad \text{Define: } \phi = 2k \\
&= \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \sqrt{r^2 + d(x^*,t)^2 + 2rd(x^*,t)\cos(2k)} dk \qquad | \quad \cos(2k) = 1 - 2\sin^2(k) \\
&= \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \sqrt{r^2 + d(x^*,t)^2 + 2rd(x^*,t)(1 - 2\sin^2(k))} dk \\
&= \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \sqrt{(r + d(x^*,t))^2 - 4rd(x^*,t)\sin^2(k)} dk \\
&= \frac{2}{\pi}(r + d(x^*,t)) \int_0^{\frac{\pi}{2}} \sqrt{1 - \frac{4rd(x^*,t)}{(r + d(x^*,t))^2} \sin^2(k)} dk \\
&= \frac{2}{\pi}(r + d(x^*,t)) \int_0^{\frac{\pi}{2}} \sqrt{1 - (\frac{2\sqrt{rd(x^*,t)}}{r + d(x^*,t)})^2 \sin^2(k)} dk \qquad | \quad \text{Following Eq. 16} \\
&= \frac{2}{\pi}(r + d(x^*,t)) E(\frac{2\sqrt{rd(x^*,t)}}{r + d(x^*,t)})
\end{aligned} \tag{17}$$

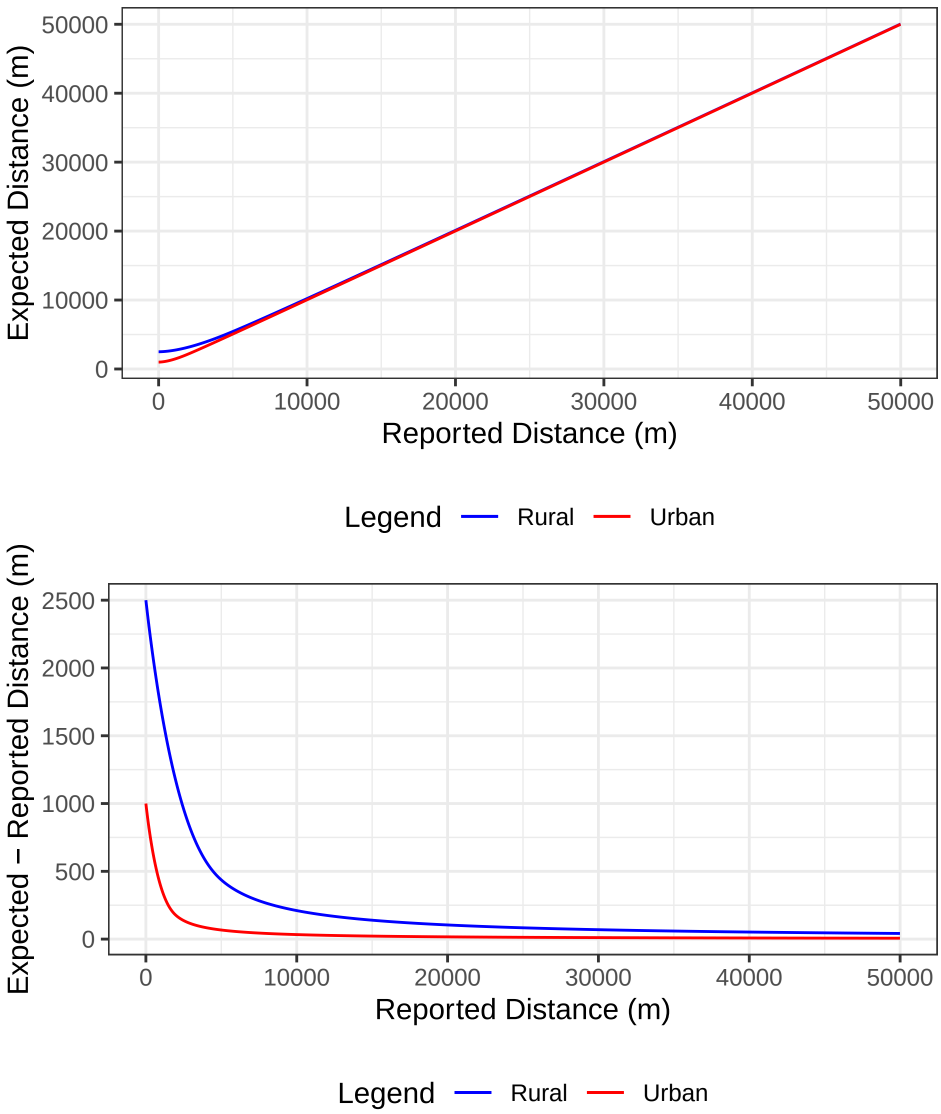


Figure A.1: Comparison: Reported and Expected Difference

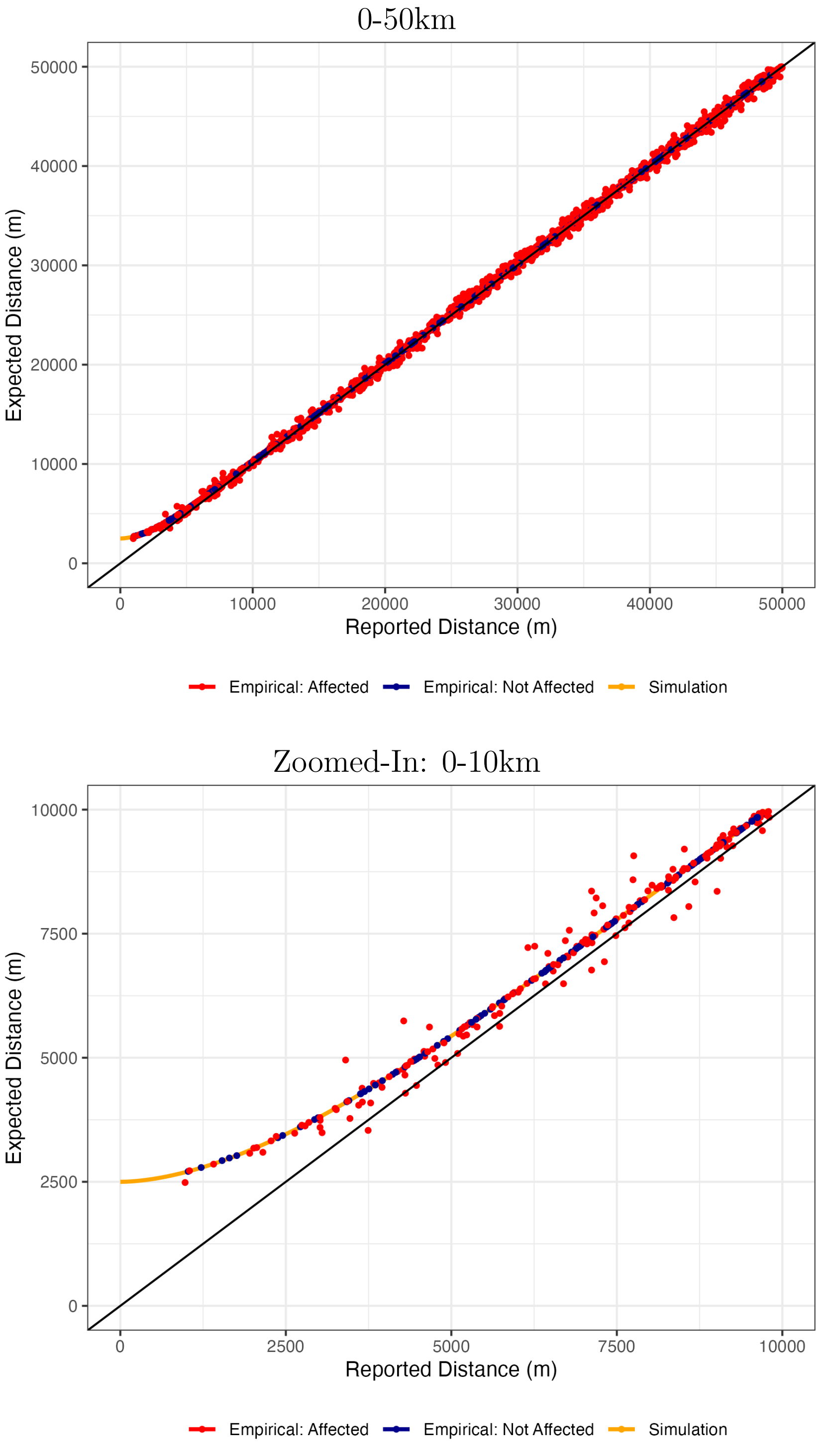


Figure A.2: Comparison: Reported and Expected Difference

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 8. 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 8. 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

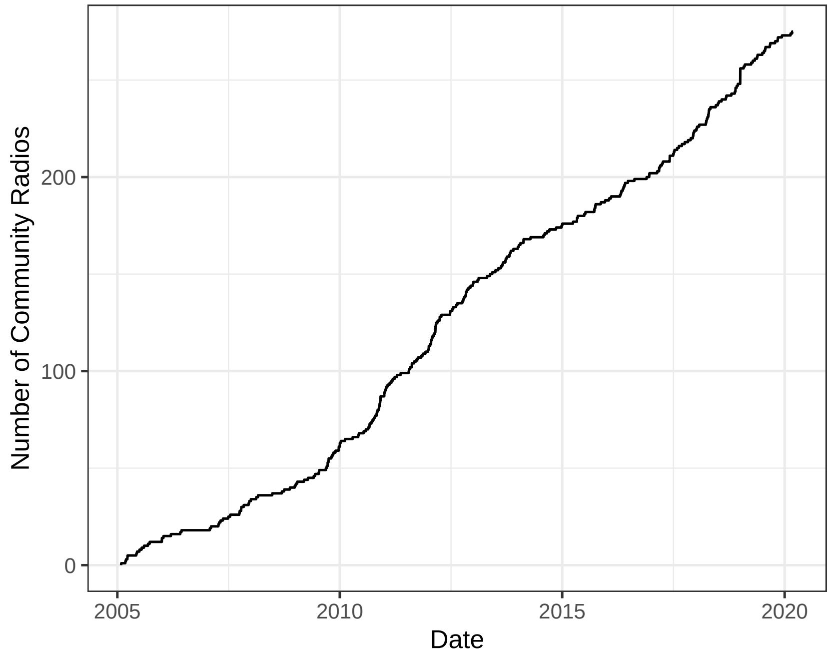


Figure B.1: Total number of CRS on air by date

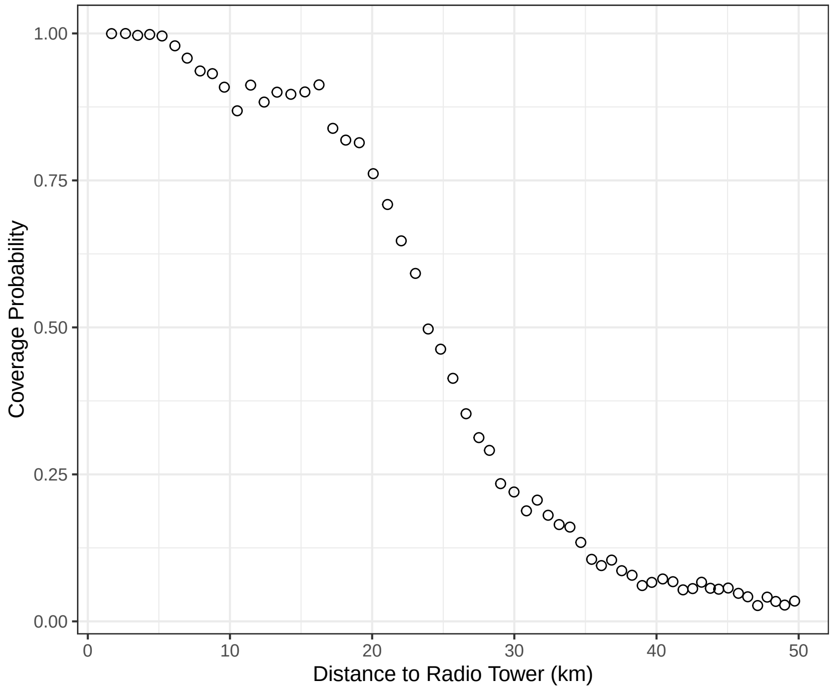


Figure B.2: Binscatter Plot of Coverage Probability and Distance to Radio

Note: The plot is created based on the *binsreg* package in R (Cattaneo et al., 2024) using default settings. The underlying data are the pooled coverage probabilities and distances to the first radio station from both the main and placebo data.

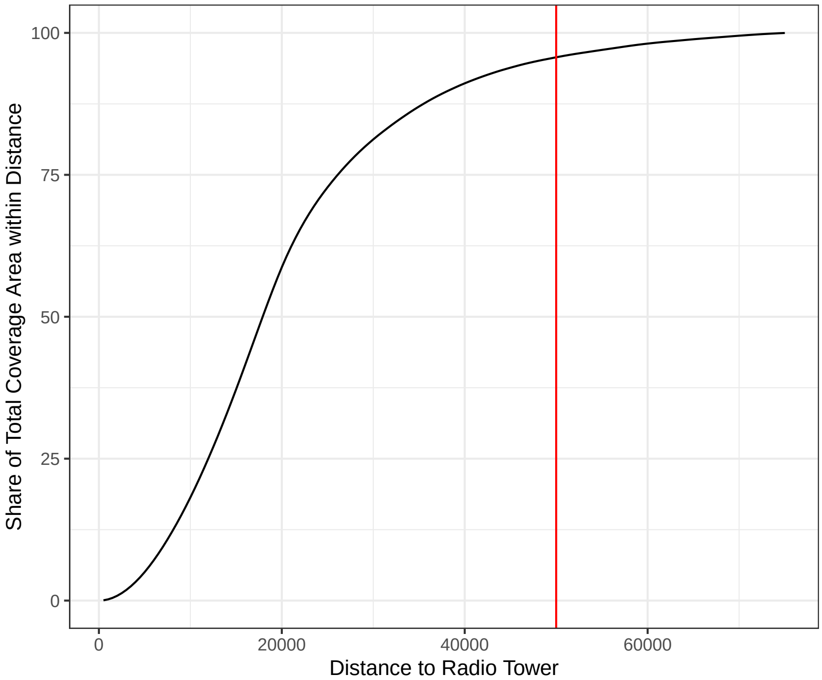


Figure B.3: Share of Total Coverage Area Within Distance

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.

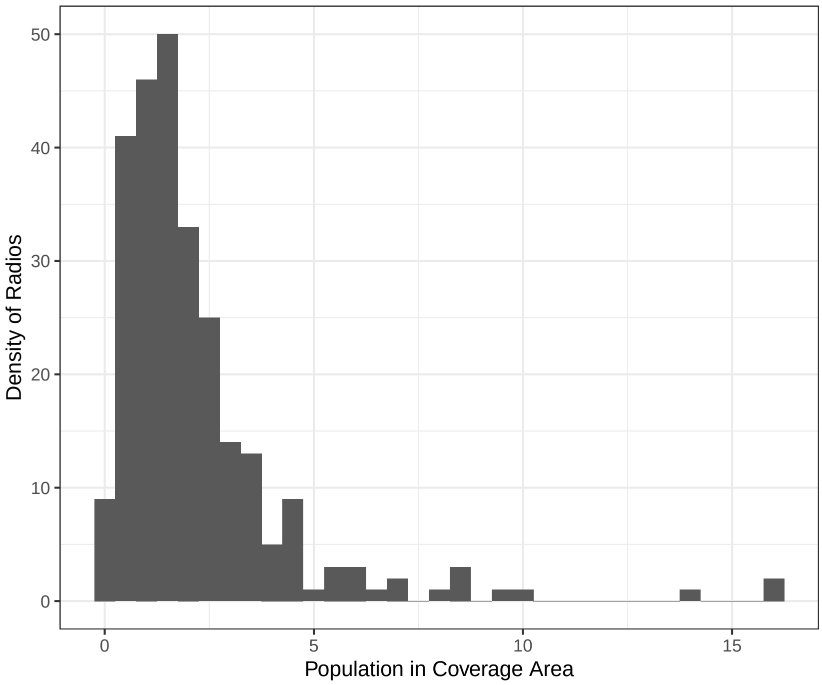


Figure B.4: Estimated Population within Coverage Area

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.

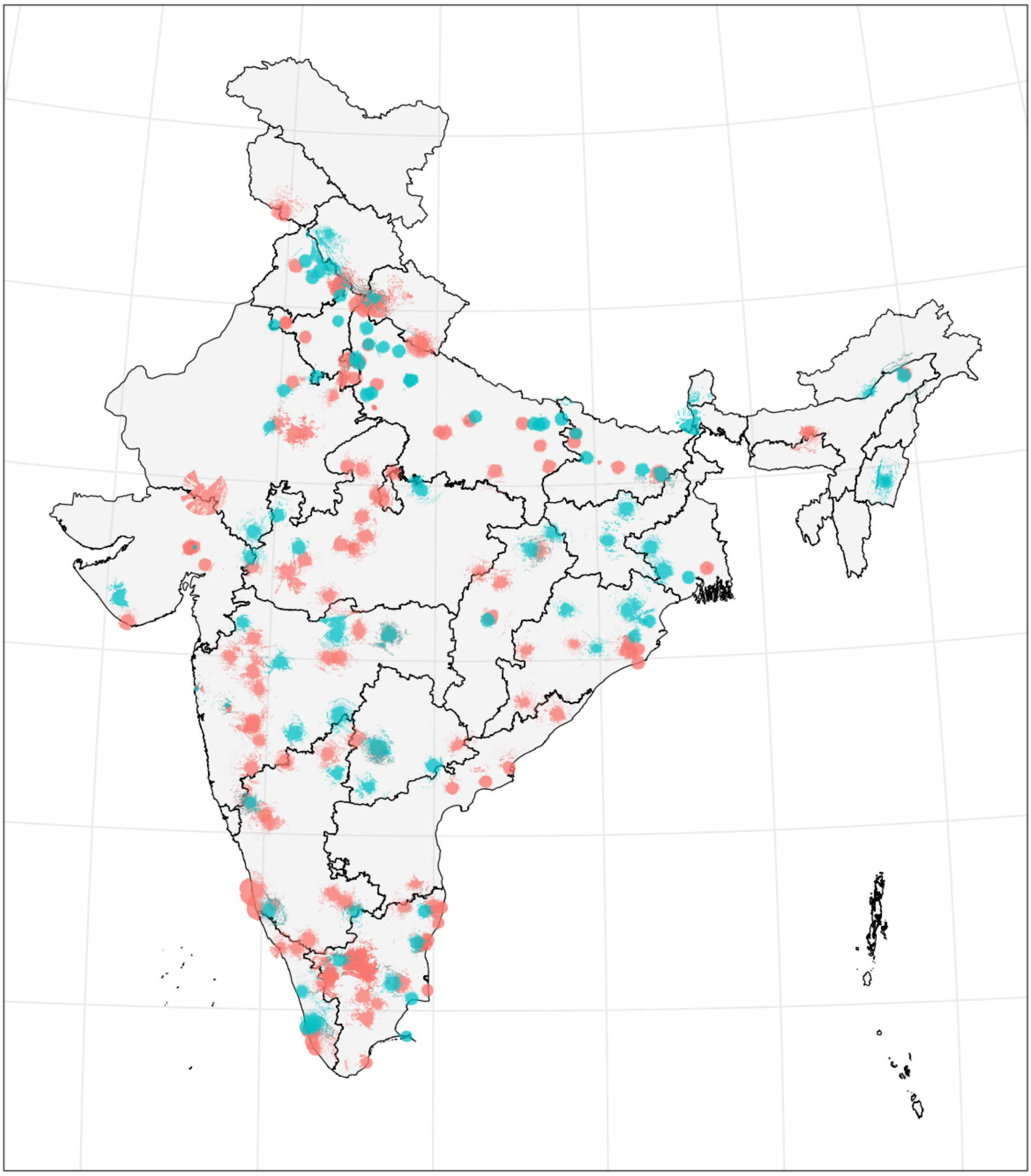


Figure B.5: Visualization of Coverage Areas of all 264 Radio Stations

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

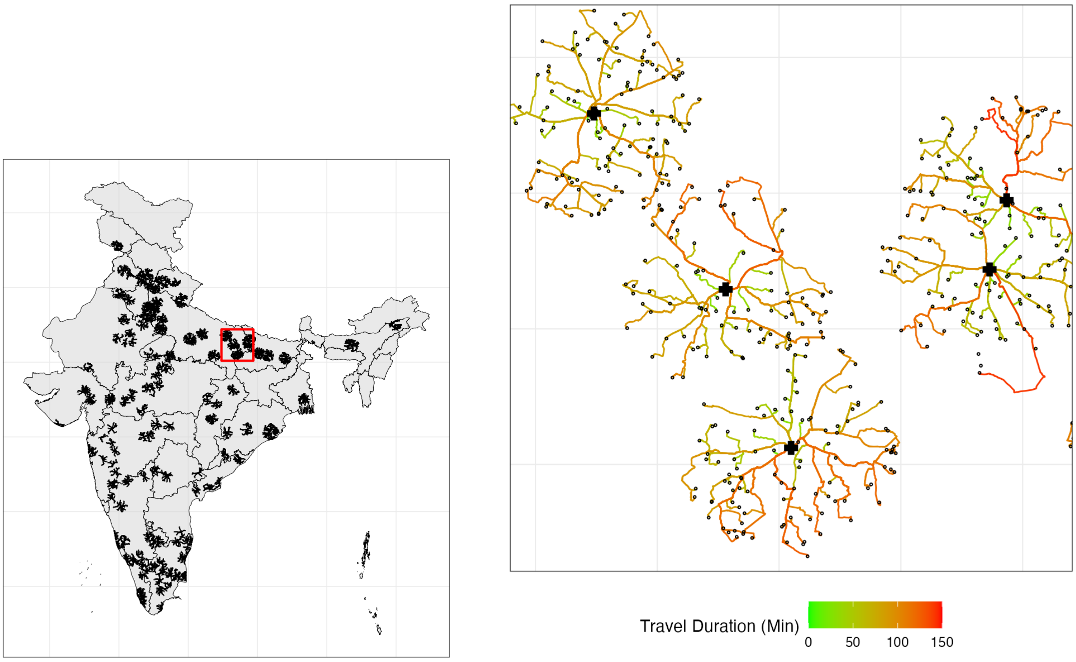


Figure B.6: Visualization of Travel Time from DHS Observation to Radio Tower

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.

C Functional Form: Explorative Analysis

In Equation 15, 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 15 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.

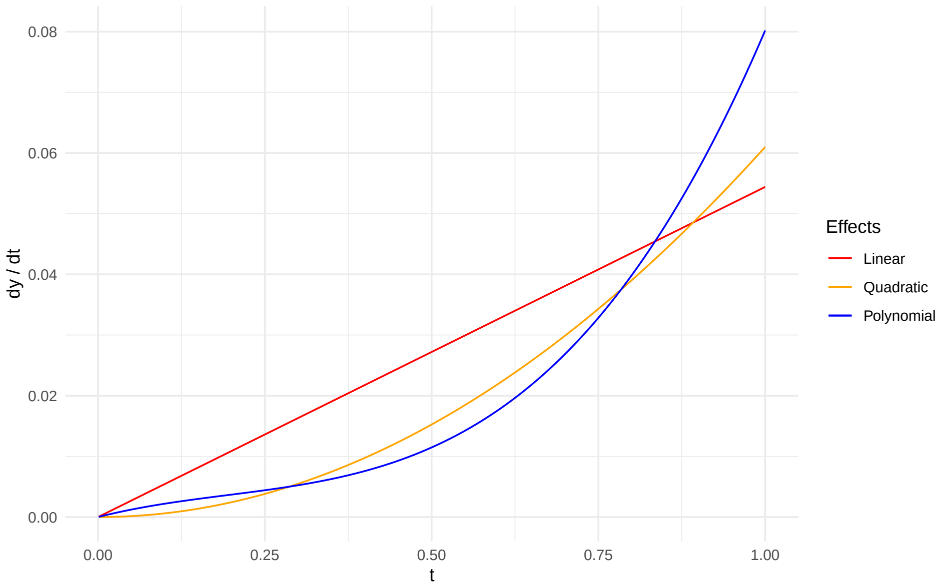


Figure C.1: Exploring Non-Linearity in Treatment Effects over Time

D Additional Results

Panel A: Linear Effects Over Time

	Radio Consumer	Radio Familiyplanning	Radio: HIV/AIDS
is female = 1 x exposure	0.033*	0.057**	0.071**
	(0.019)	(0.025)	(0.029)
is female = 0 x exposure	0.035	0.031	0.058**
	(0.036)	(0.035)	(0.029)
Num.Obs.	228 289	228 289	55 508
R2 Adj.	0.073	0.098	0.092
Mean Y	0.184	0.197	0.164
SD Y	0.388	0.398	0.37

Panel B: Quadratic Effects Over Time

	Radio Consumer	Radio Familiyplanning	Radio: HIV/AIDS
is female = 1 x exposure2	0.048**	0.066**	0.083**
	(0.020)	(0.026)	(0.036)
is female = 0 x exposure2	0.061	0.044	0.070**
	(0.040)	(0.042)	(0.029)
Num.Obs.	228 289	228 289	55 508
R2 Adj.	0.073	0.098	0.092
Mean Y	0.184	0.197	0.164
SD Y	0.388	0.398	0.37

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 5. 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 5. 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	0.038**	0.066*	−0.038**	0.009	0.010	0.019**
	(0.018)	(0.035)	(0.018)	(0.008)	(0.008)	(0.009)
Num.Obs.	196 537	196 537	196 537	196 537	196 537	196 537
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

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	0.055***	0.109**	−0.055***	0.017	0.020**	0.018**
	(0.020)	(0.045)	(0.020)	(0.013)	(0.009)	(0.009)
Num.Obs.	196 537	196 537	196 537	196 537	196 537	196 537
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

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 5. 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 5. 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)
is female = 1 x exposure	−0.014**	−0.002	−0.002	−0.039**	−0.053**
	(0.007)	(0.002)	(0.008)	(0.019)	(0.023)
is female = 0 x exposure	−0.008*	0.000	−0.006	−0.030**	−0.023
	(0.005)	(0.003)	(0.005)	(0.014)	(0.016)
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)
is female = 1 x exposure	−0.008**	0.000	−0.003	−0.029***	−0.018
	(0.004)	(0.001)	(0.005)	(0.010)	(0.015)
is female = 0 x exposure	−0.004	−0.001	−0.006	−0.021**	0.002
	(0.004)	(0.001)	(0.005)	(0.010)	(0.014)
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)
is female = 1 x exposure	−0.005	−0.001	−0.001	−0.004	−0.024*
	(0.004)	(0.001)	(0.003)	(0.013)	(0.014)
is female = 0 x exposure	−0.005*	0.000	−0.003	−0.010	−0.020
	(0.003)	(0.001)	(0.002)	(0.011)	(0.015)
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)
is female = 1 x exposure	−0.002*	0.000	0.000	0.000	−0.019*
	(0.001)	(0.000)	(0.000)	(0.003)	(0.010)
is female = 0 x exposure	0.001	0.000	0.000	0.000	0.003
	(0.001)	(0.000)	(0.000)	(0.002)	(0.008)
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)
is female = 1 x exposure	0.000	−0.001	0.002	−0.003	0.003
	(0.003)	(0.001)	(0.003)	(0.008)	(0.010)
is female = 0 x exposure	0.000	−0.001	0.002	0.006	−0.009
	(0.002)	(0.001)	(0.002)	(0.007)	(0.010)
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)
is female = 1 x exposure	0.001	0.000	0.000	0.006	−0.002
	(0.002)	(0.000)	(0.002)	(0.006)	(0.007)
is female = 0 x exposure	−0.001	0.000	0.001	−0.008	−0.001
	(0.002)	(0.000)	(0.002)	(0.005)	(0.009)
Mean Y	0.005	0	0.003	0.009	0.021

(G) Reason: Other

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
is female = 1 x exposure	0.000	0.000	0.000	−0.010	0.007
	(0.004)	(0.002)	(0.003)	(0.009)	(0.015)
is female = 0 x exposure	0.002	0.000	0.001	0.002	0.002
	(0.004)	(0.002)	(0.003)	(0.010)	(0.013)
Mean Y	0.018	0.003	0.008	0.035	0.065

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 5. 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)
is female = 1 x exposure2	−0.020***	−0.005*	−0.007	−0.036	−0.077***
	(0.008)	(0.003)	(0.007)	(0.025)	(0.024)
is female = 0 x exposure2	−0.007	−0.001	−0.006	−0.019	−0.035**
	(0.005)	(0.003)	(0.005)	(0.015)	(0.016)
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)
is female = 1 x exposure2	−0.006	0.000	−0.002	−0.018	−0.015
	(0.004)	(0.001)	(0.004)	(0.013)	(0.015)
is female = 0 x exposure2	−0.004	−0.001	−0.007	−0.021**	−0.004
	(0.003)	(0.001)	(0.004)	(0.010)	(0.015)
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)
is female = 1 x exposure2	−0.007**	−0.002	−0.003	−0.007	−0.027**
	(0.003)	(0.001)	(0.003)	(0.012)	(0.011)
is female = 0 x exposure2	−0.006**	0.000	−0.005	−0.011*	−0.018
	(0.003)	(0.001)	(0.003)	(0.006)	(0.014)
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)
is female = 1 x exposure2	−0.004***	0.000	0.000	−0.003	−0.026***
	(0.001)	(0.000)	(0.000)	(0.003)	(0.010)
is female = 0 x exposure2	0.000	0.000	0.000	−0.001	−0.001
	(0.001)	(0.000)	(0.000)	(0.002)	(0.009)
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)
is female = 1 x exposure2	−0.001	−0.001*	0.001	−0.005	−0.001
	(0.003)	(0.000)	(0.003)	(0.008)	(0.012)
is female = 0 x exposure2	0.000	−0.001	0.003	0.011	−0.014
	(0.002)	(0.001)	(0.002)	(0.007)	(0.009)
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)
is female = 1 x exposure2	0.002	0.000	0.001	0.009	0.000
	(0.002)	(0.000)	(0.002)	(0.007)	(0.007)
is female = 0 x exposure2	0.001	0.000	0.002	−0.009*	0.010
	(0.002)	(0.000)	(0.002)	(0.005)	(0.011)
Mean Y	0.005	0	0.003	0.009	0.021

(G) Reason: Other

	All (5-18)	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
is female = 1 x exposure2	−0.004	−0.002	−0.004	−0.012	−0.006
	(0.005)	(0.001)	(0.003)	(0.011)	(0.018)
is female = 0 x exposure2	0.002	0.000	0.001	0.012	−0.007
	(0.004)	(0.002)	(0.003)	(0.010)	(0.013)
Mean Y	0.018	0.003	0.008	0.035	0.065

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 5. 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)
is female = 1 x exposure	−0.015**	−0.051**	−0.025	0.009	0.005
	(0.007)	(0.020)	(0.026)	(0.012)	(0.008)
is female = 0 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

Panel B: Quadratic Effects Over Time

	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

Table D.5: Effect of CRS 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 5. 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	0.129***	0.036	−0.021	0.078
	(0.030)	(0.040)	(0.044)	(0.029)	(0.071)
Num.Obs.	24 411	5484	9572	7212	2143
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

Panel B: Quadratic Effects Over Time

	Autonomy Decision	Autonomy Decision (15-25)	Autonomy Decision (26-35)	Autonomy Decision (36-45)	Autonomy Decision (45-54)
exposure2	0.029	0.129***	0.025	−0.034	0.081
	(0.038)	(0.048)	(0.050)	(0.032)	(0.077)
Num.Obs.	24 411	5484	9572	7212	2143
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

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 5. 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	0.031	0.080*	−0.023	0.010	0.074
	(0.030)	(0.044)	(0.030)	(0.048)	(0.067)
Num.Obs.	34 324	13 451	10 386	8000	2487
R2 Adj.	0.165	0.114	0.116	0.097	0.125
Mean Y	0.506	0.368	0.545	0.637	0.667
SD Y	0.456	0.436	0.454	0.435	0.428

Panel B: Quadratic Effects Over Time

	Autonomy Mobility	Autonomy Mobility (15-25)	Autonomy Mobility (26-35)	Autonomy Mobility (36-45)	Autonomy Mobility (45-54)
exposure2	0.013	0.076	−0.028	−0.036	0.041
	(0.031)	(0.050)	(0.033)	(0.045)	(0.073)
Num.Obs.	34 324	13 451	10 386	8000	2487
R2 Adj.	0.165	0.114	0.116	0.097	0.124
Mean Y	0.506	0.368	0.545	0.637	0.667
SD Y	0.456	0.436	0.454	0.435	0.428

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 5. 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	−0.040		−0.105	
	(0.039)		(0.132)	
15-24 x exposure		−0.058		−0.129
		(0.043)		(0.152)
25-34 x exposure		−0.036		−0.127
		(0.037)		(0.129)
35-44 x exposure		−0.039		−0.099
		(0.039)		(0.121)
45-49 x exposure		−0.002		0.021
		(0.052)		(0.172)
Num.Obs.	33 442	33 442	33 442	33 442
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

Panel B: Quadratic Effects Over Time

	Attitude Any	Attitude Any	Attitude Count	Attitude Count
exposure2	−0.039		−0.095	
	(0.042)		(0.131)	
15-24 x exposure2		−0.054		−0.091
		(0.048)		(0.160)
25-34 x exposure2		−0.039		−0.124
		(0.040)		(0.128)
35-44 x exposure2		−0.032		−0.111
		(0.041)		(0.111)
45-49 x exposure2		−0.010		0.020
		(0.067)		(0.200)
Num.Obs.	33 442	33 442	33 442	33 442
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

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 5. 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	−0.040		−0.105	
	(0.039)		(0.134)	
15-24 x exposure		−0.060		−0.142
		(0.044)		(0.154)
25-34 x exposure		−0.036		−0.128
		(0.038)		(0.130)
35-44 x exposure		−0.038		−0.095
		(0.039)		(0.123)
45-49 x exposure		−0.001		0.025
		(0.052)		(0.173)
Num.Obs.	33 440	33 440	33 440	33 440
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

Panel B: Quadratic Effects Over Time

	Attitude Any	Attitude Any	Attitude Count	Attitude Count
exposure2	0.024		0.081	
	(0.045)		(0.126)	
15-24 x exposure2		0.013		0.081
		(0.057)		(0.164)
25-34 x exposure2		0.033		0.042
		(0.042)		(0.111)
35-44 x exposure2		0.025		0.126
		(0.043)		(0.122)
45-49 x exposure2		−0.031		−0.192
		(0.045)		(0.121)
Num.Obs.	30 970	30 970	30 970	30 970
R2 Adj.	0.103	0.103	0.088	0.088
Mean Y	0.303	0.303	0.696	0.696
SD Y	0.46	0.46	1.28	1.28

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 5. 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	−0.031		−0.011		0.012		0.014		−0.028	
	(0.032)		(0.031)		(0.021)		(0.028)		(0.027)	
15-24 x exposure		−0.042		−0.034		−0.014		0.029		−0.048
		(0.051)		(0.047)		(0.029)		(0.046)		(0.040)
25-34 x exposure		−0.037		−0.020		0.017		0.004		−0.041
		(0.037)		(0.037)		(0.023)		(0.031)		(0.031)
35-44 x exposure		−0.028		0.002		0.014		0.015		−0.018
		(0.035)		(0.031)		(0.020)		(0.030)		(0.032)
45-49 x exposure		−0.004		0.017		0.017		0.031		0.013
		(0.040)		(0.036)		(0.023)		(0.035)		(0.033)
Num.Obs.	18 392	18 392	18 392	18 392	18 392	18 392	18 392	18 392	18 392	18 392
R2 Adj.	0.082	0.082	0.072	0.072	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

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	−0.045		−0.007		0.014		0.012		−0.032	
	(0.039)		(0.036)		(0.024)		(0.032)		(0.029)	
15-24 x exposure2		−0.081		−0.052		−0.027		0.012		−0.073
		(0.060)		(0.058)		(0.031)		(0.059)		(0.047)
25-34 x exposure2		−0.066		−0.032		0.022		0.004		−0.066**
		(0.044)		(0.041)		(0.027)		(0.034)		(0.032)
35-44 x exposure2		−0.011		0.032		0.016		0.020		0.009
		(0.044)		(0.040)		(0.022)		(0.040)		(0.035)
45-49 x exposure2		−0.023		0.028		0.026		0.021		0.026
		(0.050)		(0.048)		(0.033)		(0.040)		(0.044)
Num.Obs.	18 392	18 392	18 392	18 392	18 392	18 392	18 392	18 392	18 392	18 392
R2 Adj.	0.082	0.082	0.072	0.072	0.024	0.024	0.039	0.039	0.069	0.070
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

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 5. 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)
is female = 1 x exposure	−0.004	0.003	0.043**	0.052**
	(0.013)	(0.010)	(0.022)	(0.021)
is female = 0 x exposure	0.000	0.007	0.024	0.030*
	(0.014)	(0.010)	(0.016)	(0.017)
Num.Obs.	91 376	62 630	31 737	32 339
R2 Adj.	0.136	0.094	0.127	0.137
Mean Y	0.904	0.922	0.799	0.648
SD Y	0.295	0.267	0.4	0.478

Panel B: Quadratic Effects Over Time

	Lower Primary (5-10)	Upper Primary (11-14)	Lower Secondary (15-16)	Higher Secondary (17-18)
is female = 1 x exposure2	−0.002	0.005	0.045	0.084***
	(0.013)	(0.011)	(0.028)	(0.023)
is female = 0 x exposure2	−0.002	0.009	0.017	0.052***
	(0.016)	(0.010)	(0.017)	(0.018)
Num.Obs.	91 376	62 630	31 737	32 339
R2 Adj.	0.136	0.094	0.127	0.137
Mean Y	0.904	0.922	0.799	0.648
SD Y	0.295	0.267	0.4	0.478

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 5. 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	−0.019	−0.006	−0.004	0.008	0.028
	(0.023)	(0.020)	(0.017)	(0.015)	(0.035)
Num.Obs.	190 157	196 537	196 537	167 111	34 324
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

Panel B: Quadratic Effects Over Time

	TV Owner	TV Consumer	Newspaper Consumer	Internet User	Mobile Phone Owner
exposure2	0.015	0.018	0.004	0.016	0.023
	(0.031)	(0.027)	(0.023)	(0.020)	(0.033)
Num.Obs.	190 157	196 537	196 537	167 111	34 324
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

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 5. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

	Caste SC/ST	Caste SC/ST	Muslim	Muslim	Urban	Urban	Log. Pop. Density	Log. Pop. Density	Travel Time (min)	Travel Time (min)	Log. Travel Time (min)	Log. Travel Time (min)	Proximity Borders (m)	Proximity Borders (m)	Travel Time Radio (min)	Travel Time Radio (min)
exposure	0.003		0.021		0.002		0.248		0.115		0.115		1139.465		4.355	
	(0.022)		(0.019)		(0.075)		(0.237)		(0.186)		(0.186)		(2895.347)		(2.887)	
exposure2		−0.029		0.002		0.015		0.267		0.089		0.089		512.614		2.686
		(0.025)		(0.022)		(0.087)		(0.255)		(0.224)		(0.224)		(2960.084)		(2.745)
Num.Obs.	167 111	167 111	167 111	167 111	171 903	171 903	171 903	171 903	171 903	171 903	171 903	171 903	171 903	171 903	171 903	171 903
R2 Adj.	0.074	0.074	0.097	0.097	0.361	0.361	0.796	0.796	0.586	0.586	0.586	0.586	0.985	0.985	0.825	0.825

Table D.13: Exogeneity check: correlation of treatment variation with other covariates of women empowerment

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 5. 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.2: 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 5. 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.3: 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 5. 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		0.005		0.007	
	(0.016)		(0.018)		(0.015)	
is female = 1 x exposure		0.025		0.015		0.007
		(0.018)		(0.019)		(0.015)
is female = 0 x exposure		−0.021		−0.005		0.007
		(0.019)		(0.019)		(0.016)
Num.Obs.	108 281	108 281	108 281	108 281	108 281	108 281
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
SD Y	0.481	0.481	0.434	0.434	0.362	0.362

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 5. 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)
is female = 1 x exposure	0.145	0.357	0.448	−0.019	0.151	0.187
	(0.268)	(0.303)	(0.340)	(0.307)	(0.268)	(0.221)
is female = 0 x exposure	−0.113	−0.232	−0.159	−0.489	−0.101	−0.196
	(0.199)	(0.345)	(0.342)	(0.333)	(0.269)	(0.220)
Num.Obs.	75 215	34 983	22 840	22 937	44 109	200 084
R2 Adj.	0.263	0.283	0.286	0.307	0.315	0.294
Mean Y	7.872	7.471	6.588	6.875	5.972	7.122
SD Y	5.276	5.313	5.39	5.395	5.353	5.379

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 5. 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	0.016	−0.066	0.010	−0.065	−0.126	−0.077
	(0.046)	(0.110)	(0.107)	(0.186)	(0.176)	(0.113)
is female = 0 x exposure	0.026	−0.025	−0.028	−0.246	−0.286*	−0.119
	(0.049)	(0.088)	(0.136)	(0.175)	(0.173)	(0.098)
Num.Obs.	75 528	51 086	25 964	36 825	141 241	320 039
R2 Adj.	0.620	0.325	0.163	0.156	0.217	0.529
Mean Y	1.633	5.82	8.234	9.503	9.356	6.873
SD Y	1.618	2.146	2.574	3.407	4.869	4.821

Panel B: degree obtained

	Primary	Secondary	Higher
is female = 1 x exposure	−0.006	−0.025	−0.019
	(0.013)	(0.018)	(0.019)
is female = 0 x exposure	−0.017	−0.023	−0.011
	(0.015)	(0.015)	(0.015)
Num.Obs.	155 556	155 556	155 556
R2 Adj.	0.155	0.148	0.124
Mean Y	0.785	0.397	0.24
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 exposure	−0.010	0.013	0.000	0.007	0.004
	(0.007)	(0.019)	(0.028)	(0.016)	(0.012)
is female = 0 x exposure	0.010	0.034*	−0.010	−0.012	0.004
	(0.008)	(0.019)	(0.025)	(0.019)	(0.015)
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)
exposure	−0.001	0.027	0.015	0.122*	0.060	0.009
	(0.006)	(0.026)	(0.058)	(0.069)	(0.068)	(0.091)
Num.Obs.	16 897	46 327	26 309	21 216	19 943	27 676
R2 Adj.	0.015	0.295	0.205	0.222	0.231	0.258
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 (45-54)
exposure	−0.016	0.032	−0.026	−0.025	−0.055
	(0.027)	(0.038)	(0.035)	(0.036)	(0.060)
Num.Obs.	19 055	4365	7436	5612	1642
R2 Adj.	0.164	0.155	0.103	0.115	0.098
Mean Y	0.642	0.505	0.65	0.711	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 5. 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	0.044	0.153	0.260*	−0.103	0.527	0.252
	(0.043)	(0.098)	(0.151)	(0.321)	(0.646)	(0.318)
is female = 0 x exposure2	−0.032	−0.069	−0.184	−0.197	0.162	−0.025
	(0.066)	(0.095)	(0.239)	(0.283)	(0.310)	(0.183)
Num.Obs.	75 528	51 086	25 964	36 825	141 241	320 039
R2 Adj.	0.620	0.325	0.164	0.155	0.217	0.529
Mean Y	1.633	5.82	8.234	9.503	9.356	6.873
SD Y	1.618	2.146	2.574	3.407	4.869	4.821

Panel B: degree obtained

	Primary	Secondary	Higher
is female = 1 x exposure2	0.023	0.026	0.042
	(0.044)	(0.046)	(0.046)
is female = 0 x exposure2	0.001	−0.004	0.011
	(0.026)	(0.029)	(0.017)
Num.Obs.	155 556	155 556	155 556
R2 Adj.	0.155	0.148	0.124
Mean Y	0.785	0.397	0.24
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	−0.004	0.064	−0.100	−0.096	0.040	−0.029
	(0.005)	(0.079)	(0.117)	(0.163)	(0.120)	(0.153)
Num.Obs.	16 897	46 327	26 309	21 216	19 943	27 676
R2 Adj.	0.015	0.295	0.205	0.221	0.231	0.258
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 (45-54)
exposure2	0.037	0.152**	−0.045	0.048	0.070
	(0.057)	(0.063)	(0.056)	(0.060)	(0.058)
Num.Obs.	19 055	4365	7436	5612	1642
R2 Adj.	0.164	0.156	0.103	0.115	0.098
Mean Y	0.642	0.505	0.65	0.711	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 5. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

G 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).[32] 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.

H 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 H.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

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

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
- 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 H.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 H.3). To get a better idea of the content of each topic, I further create wordclouds as shown in Figure H.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 H.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.

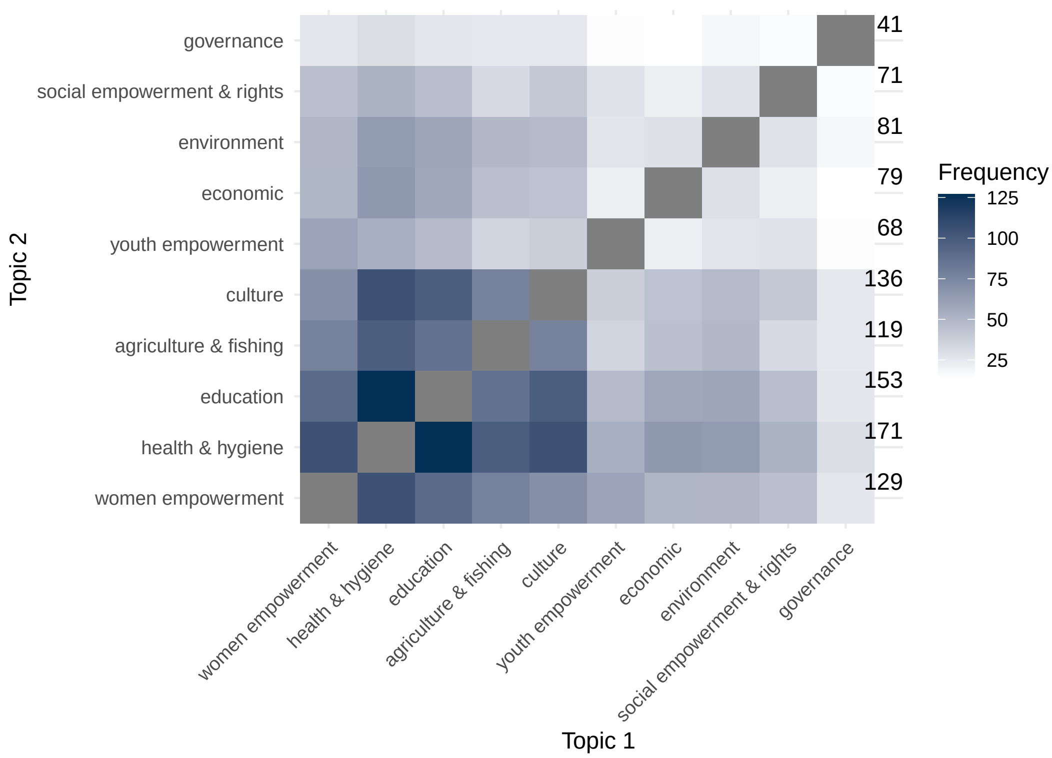


Figure H.3: Correlations in Radio Topics Based on Radio Compendia

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

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

Figure H.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.

I Radio Content II: Information from Radio Show Recordings

I.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 I.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 I.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 I.3 shows that content is uploaded in 22 languages. The website also requires radios to categorize the content uploaded. Table **??** shows the categories and subcategories of uploaded content. The table shows that health, education, inclusion and equity, and cultural development are the categories with the most radios uploading content. 100 shows by 32 radios are explicitly uploaded to the "Inclusion and Equity: Women" subcategory. However, other categories may also contain women specific content, in particular within the "Education" and "Health" categories.

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.[33] I transcribe up to 586 shows per radio. Table I.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 I.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.

[33]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 I.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).[34]

The topics are hand-labeled, building on each topic's 15 most predictive terms. These can be viewed in Tables I.4 and I.5. 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 I.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.

[34]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 I.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 I.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 I.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

Table I.4: Top 15 most predictive words by topic of LDA topic model (Part 1)

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 I.5: Top 15 most predictive words by topic of LDA topic model (Part 2)

I.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:[35]

[35] 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: 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.

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.[36]

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"

[36] 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.

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.

J Results Using Clustered Standard Errors

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 J.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 5. 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 J.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 5. Heteroscedasticity robust standard errors clustered at the subdistrict level reported in parentheses.

K 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	0.106*	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

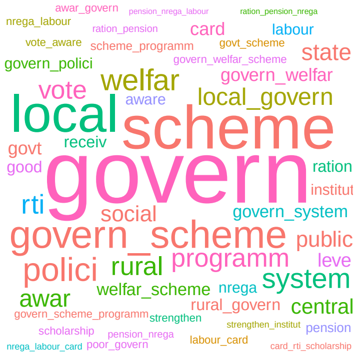
	# Children (15-18)	# Children (19-25)	# Children (26-30)	# Children (31-35)	# Children (36-40)	# Children (41-49)
exposure	0.000	−0.081*	−0.153*	−0.243**	0.018	−0.044
	(0.004)	(0.043)	(0.079)	(0.108)	(0.097)	(0.114)
Num.Obs.	15 668	43 661	25 299	20 403	19 369	27 033
R2 Adj.	0.010	0.301	0.204	0.235	0.257	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

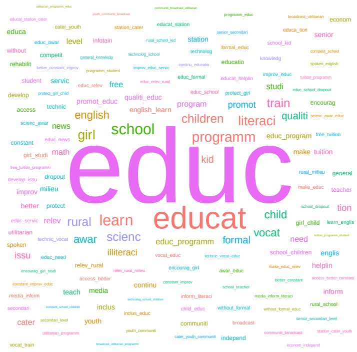
Table K.1: observations at a distance of 40km from a radio station

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 5. Standard errors in parentheses are adjusted for spatial correlation (Conley, 1999, 2010). Significance levels: *10%, **5%, ***1%.

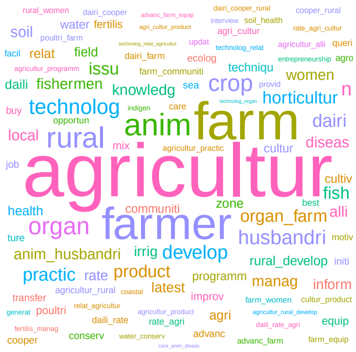


(a) Governance

(b) Health & Hygiene



(c) Education



(d) Agriculture & Fishing

(e) Culture

(f) Youth Empowerment

(g) Environment

(h) Economic

(i) Social Empowerment & Rights

Figure H.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 stemed. 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).

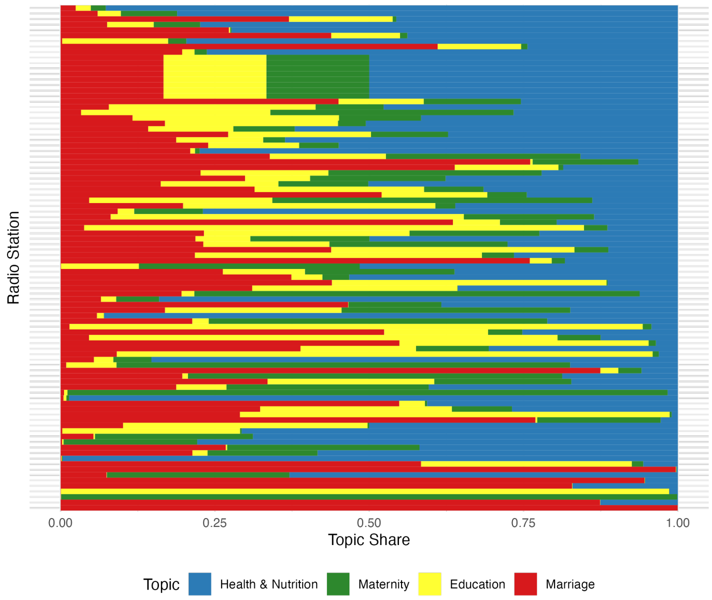


Figure I.2: Radios' Share of Content Across Women-Related Topics
Note: The radios are in the same order as in Figure 2.