# Is the Phone Mightier than the Virus? Cellphone Access and Epidemic Containment Efforts*

[Click here for the latest version]

Robert Gonzalez† Elisa M. Maffioli‡

February 3, 2022

**Abstract**

This paper examines the impact of mobile phone access on the containment of an epidemic. We study this question in the context of the 2014 Ebola Virus Disease (EVD) outbreak in Liberia. Combining novel data on cellphone towers and EVD cases, we estimate a high-resolution radio-wave propagation model that uses variations in terrain topography and the spatial distribution of cellphone towers to predict signal strength on the ground. We then employ a regression discontinuity design that compares villages at the margin of the signal strength threshold required for coverage. We find that having access to cellphone coverage leads to a 10.8 percentage point reduction in the likelihood that a village has an EVD case. Results from a novel survey collected following the epidemic suggest that this is mostly explained by cellphone access facilitating treatment provision rather than improving access to preventive care or information.

*Keywords*: Ebola Virus Disease, Mobile Phones, Technology, Information, Care

*JEL Classification*: I15, I18, O22

*We thank participants at the 2019 Econometric Society Winter Meeting, 2020 Barcelona GSE Summer Forum, PacDev 2020, the University of Michigan, UC-Santa Cruz, and University of Southern Denmark for invaluable comments and suggestions.

†School of Economics, Georgia Institute of Technology, Atlanta, GA 30313 (e-mail: robert.gonzalez@gatech.edu)

‡Health Management and Policy, University of Michigan School of Public Health, Ann Arbor, MI 48109 (e-mail: elisamaf@umich.edu)

# 1 Introduction

Infectious disease outbreaks are a major burden to low and middle-income countries (Holmes et al., 2017). For instance, we are currently in the midst of a worldwide Coronavirus epidemic and a resurgence of Ebola in the Democratic Republic of the Congo and Guinea. For this reason, assessing the effectiveness of tools that can prevent or contain these outbreaks has become a first-order policy issue. Given their widespread availability, mobile phones have the potential to be one such tool. A growing literature shows that mobile phone technology can be used to improve the delivery of health care (Braun et al., 2013; Agarwal et al., 2015; Obasola et al., 2015), to predict the spread of infectious diseases by studying mobility patterns (Bengtsson et al., 2015; Wesolowski et al., 2015), to help diagnosing diseases (D'Ambrosio et al., 2015), and as a tool for information sharing, reporting, and surveillance (Yang et al., 2009; Freifeld et al., 2010; Sacks et al., 2015). While this literature offers guidance on the design and use of specific tools that can be deployed during outbreaks, the broader question of whether access to mobile phone technology has an impact on the spread (or containment) of a disease during emergency situations remains largely unexplored.

Ex-ante, it is ambiguous whether living in an area with cellphone coverage, and thus having access to mobile phone technology, has a positive or negative impact on the spread of an infectious disease. Mobile phone technology can increase exposure to preventive care (e.g., prevention education, hygiene practices guidance), as well as facilitate access to treatment resources (e.g., reporting sick and dead people, requesting ambulances). We refer to the former as the *preventive care* channel, and the latter as the *treatment care* channel. In such cases, the likelihood of transmitting a disease is expected to be lower in cellphone coverage areas, as mobile phones can potentially lead to more desirable health behavior and/or higher relief efforts.

Cellphone coverage also enables individuals to more efficiently interact with a potentially larger network of friends and family and to improve within-network collective action during emergencies (Hampton et al., 2011, Pew Research Center, 2011, Pew Research Center, 2019, Blumenstock et al., 2016). For instance, with the advent of mobile money, cellphone access can facilitate in-network transfers to help members cope with the economic consequences of the crisis (Blumenstock et al., 2011). It is unclear whether the use of cellphone within the network would crowd out or not person-to-person interactions. However, during the early stages of an outbreak when knowledge about transmission is low and alternative government health care resources are scarce, members of a network with cellphone access can more efficiently coordinate person-to-person care of affected members of the network, reduce free-riding within the network by better monitoring the tasks of the members, or, in the event of a death, gather members for funerals.[1] An unintended consequence is then an increase in the number of person-to-person interactions within the network, which can lead to cellphone access increasing the likelihood of spreading the disease along the network. We refer to this as the *network* channel.[2]

Finally, cellphone coverage could significantly decrease the cost of access to outbreak-related information. Simultaneously, it could also open the door to misinformation and thus mitigate the returns of accessing quality information.[3] We refer to this as the *information* channel.

This paper explores the causal effect of mobile phone access –proxied by cellphone coverage– on the spread (or containment) of an infectious disease, namely the Ebola Virus Disease (EVD, hereafter) in the context of the 2014 West Africa epidemic in Liberia. We investigate whether cellphone coverage affects the likelihood that a village reports an EVD case by employing several novel sources of data. First, we use data on EVD cases compiled by the authors from primary records obtained from the Liberia's Ministry of Health (MOH). This dataset encompasses the entire set of villages in the country for the whole duration of the epidemic. Second, we gather data on the location and characteristics of cellphone towers across Liberia in the year 2013 –just prior to the outbreak– obtained from the Liberia Telecommunications Authority (LTA). Third, we explore mechanisms using original survey data conducted six months after the end of the epidemic on about 2,000 respondents across Liberia along with more than 233 million anonymous call detail records (CDR) obtained from

[1] In the case of free-riding issues, by lowering the cost of communication between members of the network, mobile phone access can enable members to repeatedly check on other members that are supposed to take care of affected members, or that have not confirmed attendance to a funeral. This, in turn, can increase person-to-person interactions between affected members and would-be free-riders within the network.

[2] Similarly, connected individuals tend to migrate and, more generally, move around more (e.g., Blumenstock (2012)). The implications within our context is that this can contribute to the spread of the disease.

[3] Refer to World Health Organization (2014); Oyeyemi et al. (2014); Onyeonoro et al. (2015); Allgaier and Svalastog (2015); Pathak et al. (2015); Roberts et al. (2017) for studies on how misinformation can disrupt epidemic response.

one of the major mobile network operators in the country.

Researchers studying the impact of mobile phone coverage on any outcome of interest face two main hurdles: how to accurately measure coverage, and how to address the endogenous selection of locations and people into coverage. This paper proposes a novel method to address these issues. First, we measure mobile coverage by estimating a high spatial resolution radio-wave propagation model widely used by regulatory agencies and businesses to model coverage, due to its high accuracy (Crabtree and Kern, 2018). This model combines cell tower footprint with information on terrain topography and more than a dozen variables (climate, terrain conductivity, antenna and transmitter characteristics, etc.) to provide a measure of signal strength at each point on the ground. To the best of our knowledge, this is the most accurate measure of mobile coverage available in the literature.[4]

Second, we employ a regression discontinuity (RD) design that uses the signal strength obtained from the propagation model as the forcing variable and the minimum signal needed for coverage as the cutoff. To account for selection into coverage, we limit our analysis to villages within a margin of this signal strength cutoff. Within this margin, whether a village receives *just* enough signal strength or not is determined by minor exogenous variations in topography that lead to arbitrary diffractions and blocking of the signal. We confirm this empirically by exploring a rich set of ex-ante village characteristics to predict coverage: topographic characteristics–not demographic or socioeconomic characteristics–are the sole predictors of coverage within a close margin of the coverage cutoff. Further analysis of these village characteristics also suggests a smooth transition across the cellphone coverage cutoff, and thus little indication that the likelihood of an EVD case is explained by these characteristics jumping at the cutoff.

We note that although the measure of coverage used in this paper is accurate, near the coverage cutoff factors such as weather conditions, call traffic, etc., may lead to day-to-day changes in signal strength. Therefore, one should interpret the estimated signal strength as

[4] Early approaches in the literature to determine coverage include using whether a given location (town, municipality, etc.) has a tower nearby (e.g., Jensen (2007); Aker (2010)), and using a fixed radius around each tower to assign whether a given location receives coverage (e.g., Shapiro and Weidmann (2015)). These approaches do not necessarily capture coverage on the ground since they inherently ignore the effect of topography on signal diffusion; thus, falling within a certain distance of a tower does not necessarily imply coverage. More recent approaches using GSMA coverage maps (e.g., Guriev et al. (2020); Manacorda and Tesei (2020); Gonzalez (2021)) provide a more accurate measure of coverage as they take into account topography; however, they do not provide a measure of signal strength within coverage areas.

coverage under typical or average conditions. This is not a major concern in our setting given that we are interested in annual coverage rather than daily coverage. Nonetheless, the effect of coverage studied in this paper should be interpreted as an intent-to-treat (ITT) given the potential fuzziness in day-to-day signal availability at the margin.

Overall, we find considerable evidence that cellphone coverage helped contain the spread of the disease. In line with graphical evidence, RD estimates show a 10.8 percentage point reduction in the likelihood that a village with just enough coverage reports an EVD case relative to villages that are just under the cutoff. Additional results using a panel-RD specification that exploits monthly variation in EVD incidence, provide significant evidence of containment effects. We find that the likelihood that a village reports an EVD case in a month, given other EVD cases within the district in the previous month, is reduced by 1.9 percentage points if the village has cellphone coverage. Villages without coverage are not as shielded, reporting instead a 1.6 percentage point increase in the likelihood of EVD given past exposure to EVD within the district.

Our findings are robust to a battery of robustness and falsification checks. First, our results do not change after controlling for "free-space" signal strength (signal strength in the absence of topography) (Olken, 2009). By comparing locations that would have received the same signal strength in the absence of topography, this exercise ensures that the identification relies solely on variation due to exogenous topographic characteristics Furthermore, our findings hold when: (i) using alternative measures of the outcome, (ii) using coverage one year after the epidemic as a falsification test, (iii) using both geographic distance and signal strength distance as forcing variables, (iv) accounting for potential non-compliance in access to cellphone coverage, and (v) assessing the sensitivity of our RD estimates to near-cutoff observations.

We explore several channels underlying the relationship between cellphone coverage and the likelihood of being affected by the epidemic. As a first step, we take advantage of the introduction of a toll-free, nationwide phone alert system established for rapid notification and response (i.e., a hotline) to provide preliminary evidence on the care and network channels. We expect the network channel to be particularly relevant during the pre-hotline period as government-provided emergency resources were scarce, forcing individuals to rely on their network for support. Similarly, we expect the preventive and treatment care channels to be more relevant after the introduction of the hotline given that they depend on the existence of a tool, such as a hotline, that can connect individuals to the appropriate agencies (ambulances, Ebola treatment units, NGOs providing educational material, etc.). We find that cellphone coverage leads to a large and significant drop in the likelihood of EVD in the period *after* the introduction of the hotline (August 2014), but not prior to this.

We explore the *preventive* and *treatment care* channels using survey data. We test whether individuals in coverage areas are more likely to receive preventive care, by asking individuals whether health workers, officials, and community task-forces came to their village to explain EVD, to hold hygiene meetings, to bring information, to teach safe burial procedures, or to bring preventive materials. We find no statistically significant evidence on this channel. Second, we test whether survey respondents in coverage areas are more likely to receive treatment care during the epidemic. We find that treatment care plays a bigger role in explaining the effect of coverage on the likelihood of an EVD case. Survey respondents in cellphone coverage areas are more likely to report that someone came to take sick people, that ambulances arrived on time, and that care centers were placed near their villages. Putting together the preventive and treatment care outcomes into summary index measures (Kling et al., 2007), we report a statistically significant effect for treatment care, but not for preventive care. Overall, the findings suggest that, while having access to mobile phones, on average, did not increase exposure to preventive care, it did help respondents report their need for relief efforts and receive more treatment resources.

We proceed by testing the *network* channel using two alternative measures of a village's network. First, we use original data on more than 233 million anonymous CDR for the universe of mobile subscribers of one of the largest mobile networks operators in Liberia. By including information on date and time of calls, identifier for cell tower where calls originated, and identifier for cell tower receiving the calls, the data allow us to create tower-to-tower measures of connectedness based on day-to-day call behavior. We define networks for the villages within the catchment areas of these connected towers. Second, we classify all villages across Liberia according to their clan using the latest available pre-outbreak (2008) census data. Clans are groups of villages that, although currently considered administrative units, correspond to historical tribal chiefdoms that were gradually fused into the state (Nyei, 2014). We then test whether EVD spreads more easily if there is a "coverage match" between an affected village and other villages within the network. In other words, we assess whether the likelihood of an EVD case in villages with cellphone access increases or decreases if an affected village within the network also has coverage. We find that a "coverage match" does not lead to any meaningful change in the likelihood of EVD spread within the network.

Finally, we test whether respondents in coverage areas were more likely to receive EVD-related information and whether they were more likely to be (mis)informed about the origin of the epidemic. We conclude that survey respondents in cellphone coverage areas are as (mis)informed as respondents living in areas with no cellphone coverage. We attribute this finding to both, the widespread availability of radio across Liberia and potentially the lack of Internet access at the time of the epidemic.

This paper fits into the economics literature that investigates the economic impact of mobile phones and other information and communication technologies (ICTs) in developing economies (Aker and Mbiti, 2010).[5] It also contributes to past research exploring the role of mobile phone technology as a tool to improve a number of health-related outcomes. Some of the outcomes studied in this literature include the management of health records and health care utilization (Agarwal et al., 2015), maternal and child health indicators (Obasola et al., 2015), the remote diagnosing of diseases (D'Ambrosio et al., 2015), the quality of care, the efficiency of services, and the capacity for monitoring (Braun et al., 2013). Our paper advances this literature in three areas. First, we present a way of measuring access to mobile phone technology at a large, country-wide scale. Previous studies typically focus on limited settings where access is determined by whether individuals or health care providers employ study-specific tools and phone applications. Our approach of accurately measuring coverage over a large area allows for analyses that can assess the effect of mobile phone interventions at a much larger scale than previous work. Second, selection into the use of mobile technology is endogenous. Thus, when comparing health-related outcomes across users and non-users of the technology, it is difficult to disentangle the effect of the mobile phone interventions from

[5]These studies explore, among other things, the effects on price dispersion (Jensen, 2007; Aker, 2010; Aker and Fafchamps, 2014), education (Aker et al., 2012; Aker and Ksoll, 2018), the role of mobile money in financial transactions (Jack and Suri, 2011, 2014), also at the time of emergencies (Blumenstock et al., 2016).

the effect of other determinants of the technology such as education or attitudes towards new technology adoption. Our empirical design addresses this underlying issue. Third, while this literature focuses on studying the effect on health-related outcomes in regular, day-to-day settings, our paper explores whether the technology is effective during a health crisis–namely, a sudden-onset epidemic–in a setting characterized by general mistrust towards local and international institutions. Our paper shows that mobile-based interventions can be effective even in such settings.

This paper also contributes directly to the strand of the literature related to mobile technology and infectious diseases. Generally, studies within this area focus on how mobile technology can be used to prevent future outbreaks (e.g., using post-outbreak mobility patterns estimated from phone usage to predict the spread of the disease (Lu et al., 2012; Bengtsson et al., 2015; Wesolowski et al., 2015)), or evaluating phones as a "participatory epidemiology" tool (e.g. using phone technology for information sharing, reporting, and tracking of cases within communities (Yang et al., 2009; Freifeld et al., 2010; Sacks et al., 2015; Feng et al., 2018)).[6] While this literature evaluates specific tools that can be deployed during outbreaks, our paper answers a more general question: whether access to mobile phones among the general population can have an impact on spreading (or containing) outbreaks. We also go beyond the evaluation exercise by exploring mechanisms that can potentially explain the relationship between cellphone access and epidemic spread (or containment). Our findings show how something as simple and ubiquitous as a mobile phone can have positive externalities on economic development by allowing communities to better access treatment health care resources in times of crisis.

This paper proceeds as follows. Section 2 describes the context. Section 3 provides details on the dataset used in the analysis. Section 4 describes the empirical models and Section 5 the results. Section 6 describes the channels of impact, while section 7 concludes.

[6]Mobile phones were also used during the 2014 West Africa Ebola epidemic to collect and share data, to create and share digital maps of the diseases, to track contacts and the spread of the disease within a community (Sacks et al., 2015), and to track health seeking behavior (Feng et al., 2018).

## 2 Background on the 2014 EVD Outbreak in Liberia

The first case of EVD in West Africa occurred in Guinea near the border with Sierra Leone and Liberia in December 2013, but EVD was not confirmed in Liberia until March 2014. After the first case was recorded in the country on March 20, 2014, the Government of Liberia (GOL) started responding to the epidemic with social mobilization, case management, treatment and surveillance, water sanitation, and hygiene activities. The MOH took the lead in managing the relief efforts supported by several international institutions such as the World Health Organization (WHO), Medicines sans Frontiers, and Samaritan's Purse. The first wave of EVD was contained very quickly and, by April 9, the last EVD case for almost two months was confirmed.

However, on May 25, 2014, a new EVD case was recorded in Lofa county near the border with Guinea. By the end of June, the disease had spread to the capital city, Monrovia. A second wave of the epidemic started, deteriorating quickly. By August 2014 the situation was out of control. The GOL urgently called on the international community for a massive response, by declaring a State of Emergency on August 6. Schools and Liberia's land borders were closed, and strict control measures, including quarantines of neighborhoods and a nightly nationwide curfew, were imposed. On August 8 the WHO declared the Ebola outbreak a "Public Health Emergency of International Concern", the highest level of international alert. By the end of that month, there was a growing awareness of the need for more decentralized control and involvement of local communities: the GOL created county-level taskforces to strengthen local coordination in the fight against EVD, and an Incident Management System (IMS) devoted exclusively to the national management of the epidemic (Nyenswah et al., 2016, Hymowitz, 2017).

As the number of EVD cases continued to rise, international funding started being poured into Liberia. The United States government committed US$319 million for the response in West Africa. Other institutions, such as the World Bank, approved an additional US$105 million, with US$52 million specifically for Liberia (World Bank, 2014). Overall, about 62 countries committed US$2.3 billion to respond to the epidemic in West Africa, including US$806 million to Liberia (White House, 2014). Over time, the GOL was able to open Community Care Centers (CCCs), Ebola Treatment Units (ETUs), and coordinate safe burials and the removal of dead bodies from communities, through teams of governmental health workers. Following an assessment of the major areas of intervention during the EVD outbreak, Kirsch et al. (2017) concluded that no single intervention stopped the epidemic, rather all interventions likely had reinforcing effects. In fact, the epidemic's turning point –September 2014– coincided with a reorganization of the response, the emergence of community leadership in control efforts, and changing beliefs and practices within the population. While in the following months the epidemic was rapidly slowing down, the GOL efforts kept securing additional funding, constructing the planned ETUs and coordinating the activities of the international partners involved. By early 2015, 31 ETUs were constructed and more than 70 CCCs opened.

In January 2015 there were fewer than 15 weekly confirmed cases with the last EVD case being reported in mid-March 2015. The country was initially declared EVD-free on May 9, 2015. However, a small number of other cases reported in July and December of the same year led to the official EVD-free declaration to take place on January 14, 2016. Along with Sierra Leone and Guinea, Liberia was among the most affected countries by EVD in West Africa. In Liberia, 10,675 confirmed, probable, or suspected cases were recorded, while the cumulative number of deaths reached 4,809–the highest number in West Africa (World Health Organization, 2016). Following the epidemic, the GOL's focus shifted from the emergency response to the strengthening of the health care system.

# 3 Data

## 3.1 Ebola Data

The data on EVD cases are primarily constructed from the patient database from the MOH containing more than 19,000 patients tested for EVD from March 2014 to July 2015. The data are widely considered to be the most comprehensive database to date, since every organization taking part in the response to the outbreak was required to report cases to the MOH (Liberian Ministry of Health, 2017). Furthermore, we supplement these data with a database from Global Community, a development organization that managed all the burials after July 2014. Since the database records the village where the person resided when suspected to have contracted EVD, we were able to manually code and match the data with the entire list of 9,686 villages in the 15 counties of Liberia.

For each village, we construct the main outcome of interest as an indicator equal to 1 if at least one (probable, confirmed, or death) case was recorded in the village during the study period (January, 2014-July, 2015). We rely on this extensive margin measure because it is less likely to suffer from measurement error than an intensive margin measure such as the number of cases. As alternative outcomes, we also explore the total number of months a village was affected by EVD and whether a village recorded a suspected EVD death. We also use the date when the blood tests were performed for individuals suspected of EVD to explore the effects across different stages of the epidemic.

### 3.2 Cellphone Coverage Data and Measures

The transmission of high-frequency radio waves between cellphone towers (transmitters) and mobile devices (receivers) is what enables the transfer of information (e.g., voice calls, SMS, etc.) in a cellphone network. Therefore, one can assess the strength of coverage at a given point on the ground by modeling how these radio waves propagate across space using a signal propagation model.

This paper uses one such model–the Irregular Terrain Model (ITM)–to determine coverage strength across Liberia. The ITM is the workhorse model used by the United States government, the Federal Communications Commission, and businesses around the world to model coverage and signal propagation. This is primarily due to its high accuracy, its ability to capture terrain topography, and its predictions repeatedly validated via on-the-ground measurements (Longley and Rice, 1968; Eppink and Kuebler, 1994; Seybold, 2005; Lazaridis et al., 2013).[7]

The model provides a measure of signal strength at a given point on the ground, taking as inputs three primary sets of information: (1) the characteristics of the transmitter or cell tower (e.g., latitude and longitude, antenna height, frequency of radio wave), (2) the charac-

[7]Refer to Crabtree and Kern (2018) for a detailed discussion of the ITM. Other propagation models specifically designed to model cellphone coverage exist. However, these were mainly designed for urban and suburban environments where obstacles to propagation come from building footprints rather than topography. We also note that the ITM has been used in related literature to measure radio coverage Adena et al. 2015; Gagliarducci et al. 2020; Armand et al. 2020 and television coverage (Olken, 2009)

teristics of the receiver (e.g., antenna height and gain, receiver sensitivity), and (3) geographic characteristics of the terrain (e.g., topography, climate, terrain conductivity). Appendix A provides a detailed discussion of the variables and parameters used in the estimation of the model.

We obtain the location of cellphone towers in the year 2013 for the two largest network providers in Liberia, MTN Lonestar and Cellcom, which accounted for 91% of all mobile subscribers during the year of study (LTA, 2014).[8] The data are obtained from the Liberia Telecommunications Authority (LTA) and provide the footprint of towers by the end of 2013, just before the start of the outbreak. Appendix Figure B1 provides a map of the towers' footprint. We combine these data with the most precise global-scale elevation data model available, the 30-meter resolution ALOS Global Digital Surface Model (Open Topography, 2017) to accurately capture the effect of topography on signal propagation (JAXA, 2016).

Figure 1 presents the model output on a map of Liberia along with the location of cellphone towers. Note that areas of Liberia with no cellphone coverage in 2013 roughly correspond to non-populated areas covered by forest.[9] Received power on the ground is measured in decibel-milliwatts (dBm) and typically ranges between -50 and -140dBm with values closer to zero representing higher signal strength. However, for ease of interpretation, our measure of coverage uses the absolute value of received power. Therefore, lower dBm levels should be interpreted as stronger coverage. For GSM networks such as the one in Liberia, sufficient coverage typically entails a signal strength below 95 dBm in absolute value (GSMA, 2019), therefore areas shaded in red in Figure 1 are receiving sufficient cellphone coverage.[10]

When we explore the causal effect of cellphone coverage on the likelihood of EVD in a

[8]Other four operators existed in the country. The biggest by market share was Comium (less than 8%), while LiberCell and Libtelco had less than 1% and WAT's share was negligible. These companies had mostly users in the capital city where phone coverage is existing for the other major companies. Note also that Cellcom is currently Orange given its acquisition by Orange in 2016.

[9]Please see the location of villages in Liberia in Maffioli (2021), Figure II; see map of Liberia on forest land at http://www.fda.gov.lr/wp-content/uploads/2014/10/forest_land2.jpg

[10]Refer to section 4.1.1 for more information on this cutoff. Additionally, one concern might be that cellphone coverage estimated from the ITM might not always correspond to actual cellphone ownership on the ground. Unfortunately, our data does not allow testing this relationship as we do not have information on cellphone ownership. Instead, we use the Demographic and Health Survey 2013 to assess the relationship between phone ownership and predicted coverage from the ITM, at the district level. Appendix Figure B2 shows that there is a positive and statistically significant correlation (0.55) between the proportion of individuals reporting owning a cellphone and the proportion of villages within each district predicted to have coverage.

village, our analysis will use a dichotomous indicator for whether the village had coverage or not in 2013.

### 3.3 Survey Data

We use novel survey data gathered about six months after the end of the epidemic (Maffioli, 2020) to explore potential mechanisms of the effects. Phone numbers from 2,265 respondents in 571 villages across all of Liberia were selected through random dialing of phone numbers. These respondents were then interviewed through a combination of an Interactive Voice Response (IVR) survey to find out about their location before the beginning of the outbreak and a mobile phone survey conducted by a local NGO. About 30% of the individuals surveyed were living in areas that did not have coverage just prior to the outbreak in 2013, so this allows us to perform analyses that compare outcomes across the pre-outbreak coverage cutoff.

### 3.4 Village Location and Census Data

We obtain GPS coordinates (latitude and longitude) of each village from the Liberia Institute of Statistics and Geo-Information Services (LISGIS). This allows matching the village location data with the spatial radio-wave propagation model in order to determine signal strength for each village in Liberia. We also obtain data on road networks to construct other determinants of EVD such as distance by road (in kilometers) to the origin point of the epidemic, and the capital city, Monrovia.

In addition, we gather data from the 2008 National Population and Housing Census (LISGIS, 2008) including information on population characteristics, such as education, household size, working status, occupation, tribe, and religion. It also includes information on housing and asset ownership, which we use to create proxies of village wealth. We aggregate census data at the village level and match it to our village-level EVD and cellphone coverage measures. We use this dataset primarily as a source of covariates for the main analysis and to assess the validity of the empirical design. Finally, we access publicly available data on various measures of village exposure to relief efforts, such as the location of CCCs.[11]

[11]See the Ebola crisis page at Humanitarian Data Exchange, https://data.humdata.org/ebola.

### 3.5 Call Detail Records (CDR) database

In order to create measures of each village's network, we obtained anonymous CDR data from Cellcom Liberia, which was the second largest mobile network operator (MNO) in the country at the time of the study. These data include information on every single voice call and SMS for each subscriber. Each record includes a unique (anonymized) device identifier, date and time of call/SMS, identifier for cell tower where call/SMS originated, and identifier for cell tower receiving the call/SMS. We were able to map 152 unique tower locations covering all counties of Liberia. We use close to 234 million CDR from the universe of Cellcom subscribers in the country–more than one million unique users–for June and July of 2015. For reference, Panel (a) of Appendix Figure B3 depicts all calls during a randomly drawn day of operation (June 15, 2015). Panel (b) depicts outgoing calls for the tower servicing Ganta city, Nimba county, as an example. The line color indicates the frequency of calls between the two locations specified by the line, and the polygons depict Voronoi cells around each tower location.

# 4 Methods

### 4.1 Regression Discontinuity Design (RD)

We estimate the effect of cellphone coverage on the spread of EVD, by employing a regression discontinuity (RD) design that uses signal strength as the forcing variable and the receiver's sensitivity threshold as the treatment cutoff. A receiver's sensitivity threshold is essentially the minimum signal strength required to be able to make a voice call or send an SMS. In familiar terms, the received power is the underlying continuous measure of the "bars" displayed on a mobile phone screen, while the receiver sensitivity threshold is the point where one goes from a single "bar" to "no-service".

Our baseline RD specification is given by the following equation:

$$EVD_i = \alpha + \beta D_i + f(\tilde{R}_i) + h(\mathbf{G}_i) + \epsilon_i \tag{1}$$

where $EVD_i$ is an indicator for whether village $i$ was affected by at least one EVD case (probable, confirmed, or death) within our study period (January, 2014-July, 2015). $\tilde{R}_i =$

$-1 \times (R_i - c)$ is the received power (measured in dBm) in village $i$ net of the receiver sensitivity cutoff $c$.[12] Values of $\tilde{R}_i$ greater than zero mean that cellphone coverage is available in village $i$, while negative values mean that the location is below the sensitivity threshold and thus no cellphone coverage is available. $D_i = \mathbb{1}\{R_i \geq c\} = \mathbb{1}\{\tilde{R}_i \geq 0\}$ is thus an indicator for whether village $i$ has coverage (i.e., received power is higher than the cutoff $c$). $f(\tilde{R}_i)$ is the RD polynomial. Our main analysis uses a local linear specification with a bandwidth $h$ around the cutoff $c$, optimally determined as in Calonico et al. (2014), and a triangular weighting kernel.[13] $h(\mathbf{G}_i)$ is a flexible polynomial in topographic characteristics such as elevation and terrain slope. This ensures that the estimated effect is the result of coverage and it is not due to changes in topography captured by the radio-wave propagation model.

We note that near the coverage cutoff factors such as weather conditions, call traffic, etc., may lead to day-to-day changes in signal strength. Therefore, one should interpret the estimated signal strength from the model as coverage under average conditions. This is not a major concern in our setting given that we are interested in annual coverage rather than daily coverage. Nonetheless, the effect of coverage studied in this paper should be interpreted as an intent-to-treat (ITT) given the potential fuzziness in day-to-day signal availability at the margin.

#### 4.1.1 Determining the Sensitivity Cutoff

We define the minimum required signal strength or sensitivity cutoff $c$ based on the industry standard for the specific network used in Liberia at the time of the study. MNOs in Liberia primarily used a GSM900 network–the most common network in the world at the time–where towers use a frequency band of 900 MHz.[14] We further confirm this with our data on tower characteristics: about 93% of the towers in Liberia used a 900MHz band (GSM900), while under 7% of the towers used 1800MHz (GSM1800). These latter towers were exclusively

[12] For convenience, we multiply the normalization by -1 simply to make positive values mean more coverage while negative values no coverage.

[13] This is the equivalent of running a weighted regression that sets $f(\tilde{R}_i) = \tilde{R}_i + D_i \times \tilde{R}_i$ where weights are obtained using a triangular kernel $K(u) = 1 - |u_i|$ when $|u_i| \leq 1$ and $K(u_i) = 0$ when $|u_i| > 1$, with $u_i = \tilde{R}_i / h$. In our estimation, we use the bias-corrected estimator proposed in Calonico et al. (2014).

[14] GSM900 was used by most of the world with the exception of North America, and some Latin American and East Asian countries (Burbank et al., 2013). More information specific to Liberia in the Liberia Telecommunications Authority (LTA) spectrum management page: `https://www.lta.gov.lr/spectrum-management`.

located within the Monrovia area and therefore are not part of our estimation sample.[15]

With this in mind, we use a cutoff of -95 dBm (i.e., $c = 95$), since this is (i) the accepted industry standard set up by the GSM Association (GSMA) for GSM900 networks,[16] and (ii) this cutoff has been repeatedly tested and validated by both the GSMA and government regulatory agencies (Razally, 2015; GSMA, 2019).[17]

We further corroborate our choice of this cutoff using three separate methods designed to detect unknown discontinuities points. First, we use the Difference in Kernels estimator described in Qiu (2011) and recommended in Porter and Yu (2015) for cases similar to our setting. This procedure compares the kernel-weighted average of our outcome $Y_i$ on the right and left sides of a set of potential cutoffs $r$. Second, we employ the maximum $R^2$ strategy, described in Card et al. (2008), that estimates a polynomial relationship between our outcome $Y_i$ and the estimated signal strength $R_i$ for several potential cutoffs $r$ but within a fixed interval $[\underline{R}, \overline{R}]$. The cutoff is chosen as the $r$ that results in the maximum $R^2$. Intuitively, if there is a discontinuity at $r$, then any specification that uses a cutoff different than $r$ is misspecified. Lastly, we use a modification of the method proposed by Spokoiny (1998) adapted to our RD setting. The method fits polynomials of our outcome $Y_i$ within gradually increasing intervals around a given point $r_0$. The cutoff is chosen as the endpoint of the maximal interval for which the residuals are "well-behaved". Intuitively, after we hit the maximal interval, the polynomial will have a hard time fitting the jump in the outcome occurring at the cutoff point. All these three methods consistently point to a value of 95dBm as the sensitivity cutoff (see Appendix C for more details).

[15]These number come from MTN–the largest MNO in Liberia at the time–which provide information on the frequency bands of their towers.

[16]More specifically, the cutoff of -95dBm is the accepted standard for phone between the head and hand (BHH) position which is the most realistic position (phone over the ear). Other cutoffs are for free-space (i.e., signal propagating in a vacuum) or browsing position (i.e., phone in hand for browsing). This latter position is more relevant for networks with data/internet access and therefore more advanced than the one used in Liberia in 2013.

[17]For instance, the GSMA and OFCOM (the UK's communication regulator) have performed separate and independent tests on the sensitivity cutoffs on BHH position. The sensitivity cutoffs for the tested devices in the GSMA study were on average -94.45 (95% CI=-(-95.04,-93.86)). In the OFCOM study, the overall sensitivity cutoffs averaged -94.46 (95% CI=(-95.10,-93.82)) with the cutoffs for lower channels of the band (925.2MHz) averaging -94.93 (95%CI=(-96.21,-93.65)) and -94.83 (95% CI=(-95.89,-93.77)) for medium channels (942.6MHz). Calculations performed by the authors based on reports (Razally, 2015; GSMA, 2019).

#### 4.1.2 Internal Validity of the RD Design

Coefficient $\beta$ in Equation (1) identifies the causal effect of cellphone coverage under the assumption that potential outcome functions $E[EVD(1)|\tilde{R}]$ and $E[EVD(0)|\tilde{R}]$ are continuous at the coverage threshold $c$, where one and zero denote assignment and non-assignment into treatment, respectively. This entails that observable and unobservable characteristics must transition smoothly across the coverage cutoff, so that villages with received power just below the cutoff can serve as a valid counterfactual for villages where coverage is just available. This is a plausible assumption within a reasonable bandwidth of analysis, as we will be comparing villages that are at the margin of cellphone coverage. At this margin, whether a village receives just enough signal strength or not is mostly determined by exogenous variations in topography. This is clearly seen in Figure 2 which compares signal strength obtained from estimating the ITM without accounting for topography (Panel a) and accounting for it (Panel b). Note that minor changes in terrain topography lead to arbitrary blocking and diffraction of the signals.

Figure 3 provides a visual depiction of our empirical strategy. In panel (a), we provide a closer look at the estimated signal strength taking as example three cellphone towers near Foya city along with the surrounding villages.[18] Panel (b) highlights villages that are part of a hypothetical RD design that uses a bandwidth of 10 dBm around the coverage cutoff.[19] First, note that at the margin of coverage, there is rich spatial variation in treatment and control villages. Most importantly, at this margin, treatment status is determined by minor changes in topography that dictate whether enough signal reaches the ground.

In order to assess how village-level characteristics change with cellphone coverage, we further explore the validity of our design by assessing various determinants of selection into coverage. Table 1 presents results from a linear regression of signal strength, measured in dBm, on a rich set of ex-ante village-level covariates.[20] If selection is not an issue around the cellphone coverage cutoff, then we should expect ex-ante demographic and economic charac-

[18]Foya is a city within Lofa county and is one of the largest cities in Liberia close to the border with Sierra Leone and Guinea, where the outbreak originated.

[19]This hypothetical bandwidth is actually quite close to the optimal bandwidth of 9 dBm used in our baseline results (section 5.1).

[20]The demographic and economic controls are obtained from the 2008 National Population and Housing Census LISGIS (2008) and thus predate the cellphone coverage outcome used.

teristics to not predict signal strength. There is evidence of significant selection into cellphone coverage when considering the entire sample of villages (column (1)): elevation, population size, and some socioeconomic indicators are strongly correlated with signal strength. Furthermore, the set of topographic, demographic, and economic controls are all jointly statistically significant. However, as we restrict our analysis to villages that are within a close window of the coverage cutoff (column (3)), only the topographic controls remain jointly significant, while demographic and economic characteristics of the villages do not explain signal strength. This suggests that, although there is significant selection into coverage when considering the entire sample, for villages at the margin of coverage, what largely determines cellphone availability are minor exogenous variations in topography, and not endogenous village characteristics. Given their importance, all baseline specifications presented in the paper control for topographic characteristics: elevation and slope.

Appendix Figure B4 further confirms that selection of villages near the coverage cutoff does not seem to be an issue as there is no significant jump in the density of the forcing variable. Appendix Table B1 also provides summary statistics for village-level characteristics for several bandwidths around the cellphone coverage threshold. Columns (1) and (2) report the mean of these variables by coverage status for the entire sample. Columns (4) and (5) repeat the exercise for villages within 20 dBm on each side of the sensitivity cutoff. Columns (7) and (8) narrow the window of analysis to a 10 dBm bandwidth. Columns (3), (6), and (9) report the clustered standard errors of the difference in means between villages with and without cellphone coverage. Comparing columns (1) and (2) confirms that, among other things, villages in areas with coverage tend to be at lower elevation and on a smoother terrain, households in those villages have a smaller average household size, higher levels of primary and secondary education, higher asset ownership and quality housing, and they live much closer to the capital Monrovia and to the closest main city. As we restrict our analysis to villages at the margin of cellphone coverage, however, most statistically significant differences disappear (columns (6) and (9)). Overall, these results provide support for the continuity assumption discussed above.[21]

[21] For a graphical depiction of the continuity of time-invariant covariates across the coverage cutoff, refer to Appendix Figure B5. Unfortunately, we do not have baseline covariates in 2013 since the main source of demographic and economic controls is the 2008 census which predate our coverage data.

#### 4.1.3 External Validity of the RD Design

Given the localized nature of the RD design, we explore how our estimation sample differs from the rest of the Liberian population to broadly characterize which sub-population our design speaks to.

Appendix Table B2 presents summary statistics for villages within our estimation sample (column 1), and outside (columns 2 and 3). We define the estimation sample for this comparison as the sample within the Calonico et al. (2014)-estimated bandwidth for our main RD design (column 1 of Table 2) at 8.9 dBm. Column 2 presents summary statistics for the sample below the lower limit of the bandwith (i.e. $< -8.9dBm$) while column 3 presents summary statistics for the sample above the upper limit of the bandwidth (i.e. $> 8.9dBm$). Broadly speaking, locations with robust signal strength (column 3) are generally closer to the towers and therefore more likely to be urban and developed, while locations far from any viable signal (column 2) are far away from the towers and therefore more likely to be isolated and rural. For ease of interpretation, Appendix Figure B6 also presents the results by standardizing all the variables in Table B2 and showing how the means of villages below (red) and above (blue) the bandwidth limits deviate from the estimation sample mean.

In terms of topographic characteristics, our sample villages are no different than villages with strong signal strength (above the bandwidth) and significantly less rugged than more isolated villages (below the bandwidth). In terms of demographic characteristics, our sample villages are quite similar to richer villages in terms of ethnicity, religion, and household size, among other characteristics. However, they are closer to more isolated villages in terms of population size, and educational attainment. In terms of economic characteristics, the sample villages tend to own less assets and have lower quality housing than villages with strong signal strength (above the bandwidth), but tend to have significantly better housing than isolated villages. They also tend to be as close to main cities and Monrovia as the richer villages and significantly closer than unconnected villages (below the bandwidth).

Overall this exercise points to clear distinctions between our sample of analysis and locations outside the sample. Relative to villages with strong signal strength (above the bandwidth), our sample villages are quite similar in terms of topographic characteristics and most demographic indicators. However, they are significantly poorer and less educated while being as centrally located as richer villages. This latter fact, in particular, points to our sample villages likely including peri-urban locations and the urban poor–a subpopulation of key policy relevance both in Liberia and the rest of the developing world. With this in mind, results from our design speak to this likely group and not to more richer/urban or more isolated/rural locations.

### 4.2 Panel-Regression Discontinuity Design (RD)

We also explore whether cellphone coverage helped contain the spread of the disease by exploiting the (monthly) time variation of the EVD epidemic. Specifically, we disaggregate our EVD measure in Equation (1) and create a village-by-month panel database. We are interested in learning whether the likelihood that EVD spreads into a village from surrounding affected villages diminishes with cellphone coverage. Our empirical specification is the following:

$$EVD_{ijt} = \alpha + \beta D_{ij} + \gamma EVD_{j(i),t-1} + \delta D_{ij} \times EVD_{j(i),t-1} + f(\tilde{R}_{ij}) + \lambda_j + \nu_t + \epsilon_{ijt} \quad (2)$$

where $EVD_{ijt}$ is an indicator for whether village $i$ in district $j$ was affected by EVD in month $t$, i.e., whether a (probable, confirmed, or death) EVD case was ever recorded in the village that month. $\tilde{R}_{ij}$, $\tilde{R}_{ij}$, and $D_{ij}$ are defined as in Equation (1) since these variables do not vary by month. $\text{EVD}_{j(i),t-1}$ is an indicator for whether district $j$, where village $i$ is located, was affected by EVD in the previous month $t-1$. $\lambda_j$ and $\nu_t$ are district and month fixed effects, respectively. The district fixed effects account for any time-invariant unobservables that may lead to endogenous selection into EVD within a village's district.

To account for endogenous selection into cellphone coverage, we integrate into our panel study a RD design that uses a linear specification in $\tilde{R}_{ij}$, while restricting our analysis to the same bandwidth as the baseline specification in Equation (1).[22] Equation (2) estimates the likelihood of an EVD case in village $i$ given that there was at least one EVD case within that village's district in the last month. Coefficients $\gamma$ and $\delta$ estimate how this contagion effect varies by whether village $i$ has coverage or not.

[22]Specifically, we let $f(\tilde{R}_{ij}) = \theta_1 \tilde{R}_{ij} + \theta_2 D_{ij} \times \tilde{R}_{ij} + \theta_3 EVD_{j(i),t-1} \times \tilde{R}_{ij} + \theta_4 D_{ij} \times EVD_{j(i),t-1} \times \tilde{R}_{ij}$.

# 5 Results

## 5.1 Graphical Analysis

Figure 4 presents regression discontinuity plots for the outcome variable $EVD_i$ in Equation (1). The solid vertical line indicates the cellphone coverage cutoff, and the signal strength is normalized so that positive (negative) dBm values represent coverage (no coverage). The circles give the averages of the outcome variable for 2 dBm signal strength bins, while the circle size is weighted by the number of villages within each bin. The solid trend lines predict $EVD_i$ using a third degree polynomial in the normalized signal strength (panels a and b) and a linear specification (panel c). The gray dots and lines provide a representation of potential $EVD_i$ by shifting upwards the predicted trend and binned averages of treated villages to the point where they intersect the observed trends in non-treated locations.

Panel (a) focuses on villages within 40 dBm of the coverage cutoff–about 68% of all villages in the sample. Overall, the likelihood of an EVD case increases as signal strength increases, and this is plausible considering that urban areas with higher cellphone coverage were the areas most affected by the epidemic. Note from the representation of potential EVD that there would have been a clear continuity in EVD likelihood in the absence of coverage. However, as soon as a village receives enough signal strength to allow cellphone use, we observe a clear drop in the likelihood of an EVD case.

Panel (b) zooms in to a window of 20dBm around the coverage cutoff. Note again, a clear continuity in our representation of potential EVD in the absence of coverage. As soon as coverage is available, the likelihood of an EVD case drops by about 10 percentage points. In addition, there is no clear change in the number of villages near the cutoff consistent with our density analysis in section 4.1.2.

Finally, panel (c) provides a graphical depiction of our baseline results presented in section 5.2 below. Specifically, the difference at the cutoff gives the RD coefficient $\beta$ in the local linear regression specification of Equation (1).

### 5.2 Main RD Estimates

Table 2 presents estimates of the RD coefficient $\beta$ in Equation (1). Given the relationship between signal strength and topography (Table 1), all specifications include controls for terrain elevation and slope. Columns (1)-(3), and (5) use a linear specification of the RD polynomial while column (4) uses a third degree polynomial in signal strength given the wider bandwidth. In column (1), we document a reduction of about 10.8 percentage points in the likelihood that a village has an EVD case relative to villages that are just under the coverage cutoff. The optimal bandwidth used in this empirical specification is about 9 dBm. The estimates remain very similar after including a set of socio-economic and demographic characteristics (column (2)), confirming that the estimated drop in the likelihood of EVD at the cellphone coverage cutoff is not explained by these covariates.

Columns (3)-(6) show that the results are robust to a set of alternative specifications. Column (3) includes a flexible polynomial in elevation and slope to capture whether the effect on EVD is simply driven by changes in topography captured by the ITM. The effect remains robust suggesting that this is unlikely. Column (4) estimates a parametric RD specification that uses almost the entire sample of villages (50 dBm bandwidth) and a flexible third degree polynomial in signal strength. The results are consistent with the optimal bandwidth estimates in columns (1)-(3), although there is a gain in precision given the larger number of observations. Column (5) confirms that the estimated effect is robust to the choice of kernel (uniform) and that are not driven entirely by observations near the cutoff. Given our binary outcome variable, column (6) estimates a Probit model within a specified bandwidth around the coverage cutoff. The marginal effect (at 7 percentage points) is not far from our previous estimates. We also probe the sensitivity of our baseline results to the choice of bandwidth. Appendix Figure B7 confirms that the coverage effect on EVD remains negative and statistically significant for a wide set of bandwidths.

Appendix Tables B3 and B4 also show that the findings are robust to two alternative measures of EVD: whether a suspected death from EVD was recorded in the village, and the total number of months the village was affected by the epidemic.[23] Lastly, it is important

[23]Unfortunately, the Ebola data do not allow us to distinguish between EVD and non-EVD deaths since all cases reported in the patient database from the MOH were suspected with EVD and not every case was tested before dying.

to highlight that if access to cellphone coverage arbitrarily led to more reporting of cases in coverage villages relative to non-coverage villages, then our estimates would be a lower bound on the actual magnitude of the drop in EVD cases due to coverage.

### 5.2.1 Panel RD Estimates

Columns (1) and (2) in Table 3 present panel-RD estimates of the effect of cellphone coverage on the likelihood that a village has an EVD case. Column (3) presents estimates of the contagion effect, namely the association between a village's district having an EVD case in the last month and the likelihood that that village subsequently has an EVD case in month $t$. Columns (4) and (5) present estimates on how this contagion effect varies by whether village $i$ has cellphone coverage or not. In line with previous findings (Table 2), we find that cellphone coverage leads to a 0.69 percentage point drop in the likelihood of an EVD case in any given month (column (1)). Column (3) provides strong evidence of contagion effects within districts: the likelihood of a village reporting an EVD case in a given month increases by 0.56 percentage points if there was at least one EVD case in the previous month within that village's district. Column (4) disaggregates this contagion effect by whether a village has cellphone coverage or not. The results provide evidence that the spread of the disease is considerably undermined by the presence of cellphone coverage. In fact, the likelihood of reporting an EVD case given past EVD cases within the district increases by about 1.6 percentage points in villages without coverage, while it significantly decreases by about 1.9 percentage points if a village has coverage. The estimates are quantitatively similar after adding controls (column (5)).

## 5.3 Robustness Checks

### 5.3.1 "Free-Space" ITM

We proceed by presenting further evidence supporting our main results. We use the ITM to estimate signal strength without accounting for topography. This is typically called a "Free-Space" model and it assumes that there is a direct line-of-sight between the tower and the receiver (Olken, 2009). Figure 2 compares the output from the two models. Note that minor changes in terrain topography lead to arbitrary blocking or diffraction of the signals.

Following Olken (2009), we use the free-space predicted signal strength and coverage as additional controls in Equation 1. This allows identifying our coverage effect using the variation in signal strength that is solely due to idiosyncratic changes in topography. Intuitively, by controlling for free-space signal strength and topography, we are comparing villages that are observationally identical in terms of their average topography and the signal strength they would have received in "free-space", but that differ in their actual signal strength due to minor exogenous changes in terrain. We estimate an expanded version of Equation (1):

$$EVD_i = \alpha + \beta D_i + f(\tilde{R}_i) + \delta D_i^{Free} + f(\tilde{R}_i^{Free}) + h(\mathbf{G}_i) + \epsilon_i \tag{3}$$

where $\tilde{R}_i^{Free} = R_i^{Free} - c$ is the received power (measured in dBm) using the "free-space" ITM in village $i$ net of the receiver sensitivity cutoff $c$. $D_i^{Free} = \mathbb{1}\{R_i^{Free} \geq c\} = \mathbb{1}\{\tilde{R}_i^{Free} \geq 0\}$ is an indicator for whether village $i$ has coverage (i.e., received power is higher than the cutoff $c$). Coefficient $\beta$ gives the effect of coverage after controlling for free-space coverage $D_i^{Free}$ and a flexible form of free-space signal strength $f(\tilde{R}_i^{Free})$. We estimate Equation (3) within a bandwidth of the coverage cutoff $c$ that is optimally determined following Calonico et al. (2014). The remaining terms are defined as in Equation (1). Note from Equation (3) that we are combining our RD strategy comparing villages at the margin of coverage with the free-space strategy. This allows obtaining results that are potentially more robust to any concerns of endogenous selection into coverage.

Columns (1)-(4) of Table 4 present the estimates for different versions of Equation (3). All specifications include topographic, economic, and demographic controls.[24] The effect of coverage remains statistically significant and similar in magnitude to our main results in Table 2 even after controlling for free-space coverage (column (1)) and free-space signal strength (column (2)). Columns (3) and (4) present the fully specified version of Equation (3). Again, we find that our coverage effect remains robust even after including a flexible polynomial in topography (column (4)).

We perform two falsification tests designed to rule out the possibility of the modeling details of the ITM driving our results. First, column (5) presents results from estimating Equation (1) using free-space coverage instead of actual coverage. As expected, we find no

[24] Refer to the notes on Table 4 for a list of these controls.

effect when using the free-space measure of coverage. Second, we estimate signal strength around towers built in 2015 when the epidemic was practically over. We limit the analysis to districts that did not have any coverage in 2013 so that we do not pick up any effect from the coverage footprint in 2013. In all, that leaves us with 383 villages, of which 77, that did not have coverage in 2013, obtained coverage from the towers built in 2015. Given the low number of observations, we cannot employ a RD design that limits observations within a bandwidth. Therefore, we run a regression of EVD likelihood in 2013 on the measure of coverage in 2015 for all villages within districts that did not have any coverage in 2013. We should find no effect as this is essentially a placebo test that uses a boundary that did not exist in 2013. Column (6) presents no statistically significant effect of 2015 coverage on the likelihood of EVD in the previous years.

#### 5.3.2 RD Estimates using Geographic and Technological Distance

The results in section 5.2 compare villages within a close window of the signal strength cutoff. This section adds geographic distance as another dimension by which villages are compared. Specifically, we construct the geographic distance from each village to its closest point on the two-dimensional coverage boundary. We then restrict the analysis to villages that are both, technologically (i.e., within a close bandwidth of the signal strength cutoff) and physically (i.e., within a close bandwidth of the spatial coverage boundary) close to each other.

Appendix Table B5 presents the results for the baseline RD (columns 1-3) and the RD specification that adds controls (columns 4-6). Columns (1) and (4) present estimates without putting any restrictions on physical distance to the coverage boundary and thus replicate columns (1) and (2) of Table 2. Instead, columns 2 and 5, and columns 3 and 4, consider a physical distance of 2 and 4 km, respectively. The results remain quantitatively similar to the main results (columns 1 and 4) after restricting the analysis to villages that are geographically close to each other. This is not surprising since there is a significant correlation between geographic and technological distance, i.e., villages close to each other in terms of signal strength tend to be geographically close.

#### 5.3.3 Sensitivity of Results to Near-Cutoff Observations

Appendix Figure B8 examines the sensitivity of our main results to observations near the signal strength cutoff. We start by performing 250 replications where, within each replication, we randomly drop 5% of the observations within a 1 dBm window of the cutoff and estimate our model (Equation (1)) using the restricted sample. We then repeat this exercise dropping 10, 15, ..., 95% of the sample within the 1 dBm window. Appendix Figure B8 plots the set of 250 estimated RD coefficients for each of the dropped sample levels (5 to 95%) along with 95% confidence intervals for each set of estimates. As we drop a larger share of the sample, the estimates tend to get smaller in magnitude relative to our baseline RD estimate presented in column (1) of Table 2. However, the estimates are consistently negative and significantly different from zero even after dropping up to 95% of the observations near to the cutoff. This suggests that, although observations nearest the cutoff understandably influence our results to some degree, they are not the sole drivers of the negative effect of coverage on the likelihood of having EVD.

#### 5.3.4 "Walking" to Coverage Boundary

Recall that our analysis estimates intention-to-treat (ITT) effects. In principle, individuals in a non-coverage village can own phones and travel to nearby coverage villages to take advantage of the technology. If that is the case, then our main estimates would be attenuated since villages assigned to control (non-coverage) areas would arguably receive some degree of treatment (coverage).[25]

For all non-coverage villages, we calculate travel distance to the nearest coverage village. We rely on the Open Source Routing Machine (OSRM) engine for calculations and OpenStreetMap as the base map.[26] Appendix Table B6 presents the average calculated travel distances. We were able to estimate travel distances for about 59% of the full sample of non-coverage villages (7,830 villages). Since the calculated travel distances for 2,816 villages (36%) came out shorter than the Euclidean distance, we drop these observations for this

[25] Note that this is a problem only if non-coverage villages that are physically near coverage villages also happen to fall within the signal strength bandwidth used in the analysis. However, since physical distance and signal strength across villages are correlated, this can still attenuate our main estimates.

[26] We use "georoute" package (Weber and Péclat, 2017) for travel distance calculations.

analysis.[27] Overall, we find that the average travel distance from non-coverage villages to the nearest coverage village is about 26 kilometers (column 1). When we reduce our analysis to villages closer to our main estimation sample–i.e., within 10dBm of signal strength threshold (column 3)–the average distance is reduced to about 5.7 kilometers with a standard deviation of about 10 kilometers.

We cannot directly observe whether individuals in non-coverage villages traveled to nearby coverage villages. However, we can assess how our baseline RD estimates respond to restricting the comparison sample to non-coverage villages spatially close to coverage villages, where individuals could have traveled to make a phone call. By restricting the sample to non-coverage villages within 2, 4, 6, 8, and 10 kilometers of a coverage village, we find no evidence that that being spatially close to a coverage village significantly attenuates our estimates (Appendix Table B7).

The fact that we find little evidence of people "walking" to coverage is contextually plausible for several reasons. First, there were significant restrictions on mobility. All affected counties were under full quarantine and strict curfews at the height of the epidemic (IFRC, 2014). Furthermore, villagers informally enforced strict protocols to limit outsiders from entering their villages (Ruble, 2015). Second, mobile phones are a technology that is inherently dependent on a network effect. If other members of the network do not have access, then there is little incentive for a single member to select into the technology (e.g., incur costs of purchasing handset, call/data plan, SIM cards, etc.). To explore this possibility, Appendix Figure B9 presents a histogram using 5 percentage point bins of the share of villages within a clan that have access to coverage. Clans can serve as a proxy for a village's network based on historical links between villages within clans. A key takeaway is that in most clans (about 63%) only a small fraction of villages actually have coverage (less than 5%). This suggests that individuals within these clans, even if they live near a coverage village, might be less willing to incur the costs of the technology until a larger mass of villages within their own clans receives coverage. Lastly, financial uncertainties inherent to epidemics can lead individuals in non-coverage villages to delay investments in the technology (handset, call/data

[27] The calculated travel distance can be shorter than the Euclidean distance if a village is relatively far from, or not directly connected to, the road network used in the calculation. We use the best available road segments but this leads to an arbitrarily short travel distance for part of the sample. Also note that the travel distances for 375 non-coverage villages (5%) could not be calculated.

plans, SIM cards, etc.).

## 6 Channels of Impact

This section explores several potential channels underlying the relationship between cellphone coverage and the likelihood of a village being affected by the epidemic.

Cellphone access can better connect individuals to preventive care resources (e.g., prevention education, hygiene practices guidance), as well as facilitate access to treatment resources (e.g., taking sick and dead people, ambulances). We refer to the former as the *preventive care* channel and the latter as the *treatment care* channel.

Cellphone coverage also enables individuals to more efficiently interact with a potentially larger core network of friends and family and therefore improve within-network collective action during emergencies (Hampton et al., 2011, Pew Research Center, 2011, Pew Research Center, 2019, Blumenstock et al., 2016). For example, individuals living in a cellphone coverage area can more easily tap into their network if they need care, transfers, or if they want to gather family for events such as funerals, one of the main factors contributing to how quickly EVD was transmitted (Alexander et al., 2015, Fallah et al., 2015). As a result, within-network person-to-person interactions–potentially with sick (or dead) members of the network–may increase. This is especially concerning during the early stages of the epidemic when knowledge about transmission was low and government-provided alternatives to in-network care were scarce. We refer to this mechanism as the *network* channel.

Last, cellphone coverage can significantly decrease the cost of access to information. This in turn, can increase exposure to outbreak-related information (e.g., latest recommendations, preventive measures, treatment resources, etc.). Cellphone coverage also considerably reduces the cost of access to misinformation potentially damping any benefits from increased access to quality information.[28] We refer to this mechanism as the *information* channel.

Overall, if cellphones are not perfect substitutes for in-person interactions–especially during an emergency, we might expect the network channel to increase the likelihood of EVD

[28] The spread of misinformation during epidemics is significant (Venkatraman et al., 2016; Ortiz-Martínez and Jiménez-Arcia, 2017; Carey et al., 2020). The 2014 West Africa Ebola epidemic was no exception (Krishna and Thompson, 2019). Refer to World Health Organization (2014); Oyeyemi et al. (2014); Onyeonoro et al. (2015); Allgaier and Svalastog (2015); Pathak et al. (2015); Roberts et al. (2017) for information on how misinformation can disrupt epidemic response.

(spread) in cellphone coverage areas, while the preventive and treatment care channels to decrease it (containment). The effect of information on containment or spread will depend on the type of information provided (quality information *versus* misinformation).

It is also possible that access to cellphones generated benefits pre-pandemic for individuals living in coverage areas (e.g. improvements in market efficiency (Jensen, 2007; Aker, 2010), literacy (Aker et al., 2012; Aker and Ksoll, 2018), access to mobile banking (Jack et al., 2010; Jack and Suri, 2011), and risk sharing (Jack and Suri, 2014; Blumenstock et al., 2016)). This may enable individuals in coverage areas to better cope with the epidemic shock and thus leading potentially to lower incidence within these areas. Yet, our evidence suggests that households in villages at the margin of coverage (Appendix Table B1) as well as individuals in the survey sample (Appendix Table B8) are similar across the signal strength cutoff in terms of socio-demographic characteristics including wealth, job opportunities, and education. One exception is the education level in the survey sample which is slightly higher in cellphone coverage areas. However, results are robust to controlling for this covariate (not shown). Overall, this suggests that the income channel do not play a major role in reducing EVD likelihood at the cellphone coverage margin.

### 6.1 Preliminary Evidence

Recall that we document a negative effect of cellphone coverage on the likelihood that a village reports an EVD case (Tables 2 and 3, column (1)). These results then suggest that the network channel is either trivial, or that the care channels dominate whatever detrimental effect cellphone access might have –via the network channel– on the spread of the disease.[29] We will discussed the information channel separately later.

We use the introduction of a hotline set up during the epidemic as a first step in disentangling the relative importance of the network and care channels. Specifically, the hotline was a toll-free, nationwide phone alert system established for rapid notification and response, in collaboration with private cellular telephone companies (Kirsch et al., 2017), thought as a vital link between the public and the government-provided relief efforts. The GOL General Service Agency opened a call center on August 7, 2014 to answer callers' questions about

[29]Note that in the presence of a non-trivial network channel, our estimates would have a downward bias.

EVD, and to enter requests to dispatch ambulances to take sick individuals to treatment centers, or to dispatch management teams to pick up suspected corpses for safe disposal.[30]

The success of the care channels likely depends on the existence of this tool (e.g., a hotline) that can connect individuals to the appropriate agencies (ambulances, Ebola Treatment Units, NGOs providing educational material, etc.). Therefore, we argue that the preventive and treatment care channels are relevant *after* the introduction of the hotline. Consequently, any effect of cellphone coverage during the pre-hotline period is likely attributed to the network channel, given that alternative, government-provided emergency resources were scarce and inefficient at that point in time and individuals were likely relying on their network for relief during this period. We should then expect the effect of cellphone coverage to be much larger after the introduction of the hotline.

Table 5 provides estimates of Equation (1) separating the analysis by whether the hotline was in place or not. Note that the effect of coverage in the pre-hotline period is indistinguishable from zero. After the introduction of the hotline, we document a consistent and robust drop in the likelihood that a village has an EVD case. It is important to highlight that if the hotline arbitrarily led to more reporting of cases in coverage villages then the results in Table 5 would underestimate the magnitude of the drop in EVD cases in the post-hotline period. Appendix Figure B10 confirms these effects for a wide set of bandwidths.

Note that the percentage of villages with any EVD case in the period pre-hotline is much smaller than the percentage of villages with any EVD case in the period post-hotline, so we cannot fully exclude that the lack of statistically significant results pre-hotline might be partially explained by the limited variation in the data. Despite this caveat these preliminary findings point to two main takeaways. First, we find suggestive evidence that the network channel is less important in the pre-hotline period. Second, we find suggestive evidence of the care channels likely at play given how EVD responds following the introduction of the primary tool (hotline) used to implement these channels.

[30]While the center received queries from nearly 1,000 people within its first two days (Kirsch et al., 2017), pranksters' calls were also common especially at the beginning (Baker, 2014). The number of ambulances to be dispatched to collect suspected cases was also limited compared to the volume of calls received. It took few months to be fully effective.

### 6.2 Preventive and Treatment Care Channels

In order to explore the preventive and treatment care channels, we use the novel survey data described in Maffioli (2020). However, note that the survey sample is not representative of the national population, instead it is biased towards respondents with access to a mobile phone during the time of the survey, i.e. male and educated individuals from urban areas.[31] However, once we restrict the analysis to a small bandwidth around the cellphone coverage cutoff to assess the continuity of a number of individual-level characteristics in the survey, we find that most differences disappear (Appendix Table B8). This gives us confidence to implement a RD design similar to the one used in our main results.[32]

Using the survey data, we construct self-reported measures of access to *preventive care* and *treatment care*. Consider that we do not directly observe hotline-specific call behavior. However, we can use a set of outcomes that capture whether cellphone access allowed communities to be more exposed to either care efforts. This can happen if, for instance, cellphone access improves coordination between communities and teams in charge of delivering preventive or treatment care. The survey asked whether during the Ebola crisis anyone from the government, health workers, NGO, or international organizations came to the community, and if so, what was the purpose of the visit: explain what EVD was, hold hygiene meetings, distribute prevention material, do contact tracing, explain how to conduct safe burials, take sick people or dead bodies. Respondents were also asked directly whether a taskforce came to their villages as the taskforce was directly set-up to bring information about EVD. In addition, we asked respondents to report how long it took, on average, for ambulances to get to their communities, whether the ambulance came very late, or never came, after the hotline was called for people sick with Ebola symptoms. Finally, we use publicly available data on the location of CCCs to construct a variable equal to 1 if the village had a CCC within a 10km radius.

Specifically, for the *preventive care* channel (Table 6), we study whether individuals,

[31] We refer to Maffioli (2020) for more details on the methodology used to sample and screen respondents and to gather data, and for more details on the sample characteristics and how it compares to a nationally representative sample.

[32] Note that using a bandwidth of 10 dBm (columns (7)-(9)) there is still a slight difference in education levels and in Kpelle tribe. Given this, we make sure to include these two variables in our set of covariates in all the analysis implemented using the survey data.

among those who report any response during the Ebola crisis, report that someone (from government, health workers, local or international NGOs) came to their village to do contact tracing (column (1), explain what EVD was (column (2)), to teach EVD-related hygiene practices (column (3)), whether a community taskforce came to share information on EVD and how to prevent it (column (4)), or someone came to explain how to conduct safe burials (column (5)) and brought preventive material such as buckets for chlorine (column 6). For the *treatment care* channel (Table 7), we explore indicators for whether someone came to take sick people or dead bodies (columns (1)-(2)). We further examine if individuals report that ambulances arrived within 4 hours after being requested through the hotline (column (3)), and whether a CCC was built within 10 kilometers of the village (column (4)).[33] We replicate the RD design and estimate Equation (1) using each of the information and care measures as our outcomes of interest.[34]

Table 6 presents results for the *preventive care* channel. Respondents in coverage areas are 10.2 percentage points more likely to report contact tracing (column (1)), 12.3 percentage points more likely to report that someone came to their village to explain what EVD was (column (2)), 13.5 percentage points more likely to report that the community taskforce came to teach preventive measures (column (4)), 9 percentage points more likely to report that someone came to explain how to conduct safe burials (column (5)), and 7.9 percentage points more likely to report that someone brought preventive material (column (6)). Instead, respondents in coverage areas are less likely to report that hygiene meetings were held in their villages (column (3)). This might be consistent with the fact that in areas with cellphone coverage simple information on hygiene practices, such as washing or not shaking hands, can be channeled through mobile phones or radio, instead of using and sending already limited personnel. On the other hand, even in areas with cellphone coverage, it might still be necessary to send health workers to explain and show how to conduct safe burials or to gain the trust of individuals with limited knowledge on EVD. Overall, we find that, while the signs of the coefficients are, generally, in the right direction, the results lack statistical power.

[33] In the survey data, 4 hours is the median time taken from an ambulance to reach the village of destination from the dispatch.

[34] The analysis uses village clustered standard errors given that the sample selection in the survey data was done at the individual level, and outcomes could be correlated for individuals within the same village.

Table 7 explores the *treatment care* channel. We find evidence that survey respondents in coverage areas are 22.3 percentage points more likely to report that someone came to take sick people (column (1)). They are also 20.6 percentage points more likely to report that when they called an ambulance, it arrived on time (column (3)), and significantly more likely to report that a CCC was placed near their village (column (4)).[35] We find a null effect on the likelihood of someone coming to take dead bodies.

Combining the results in Tables 6 and 7, we conclude that the treatment care channel seems to play a bigger role in explaining the effect of coverage on the likelihood of an EVD case. In the case of the preventive care outcomes, while the signs of the cellphone coverage effect are generally reasonable, the results lack precision. This is further corroborated when we explore summary indexes of each of the two care channels constructed following Kling et al. (2007). The index of all preventive care outcomes results in a null coverage effect (Table 6, column (7)), while we find a positive and statistically significant effect in the case of the combined treatment care outcomes (Table 7, column (5)).[36] Overall, these results are quite plausible given our setting. In the midst of a crisis, an additional ambulance or a CCC near a village is likely more impactful and immediately observed, compared to an additional information session on prevention. The effects of preventive care may take longer to materialize as they essentially entail a change in health behavior.

### 6.3 Network Channel

We are unable to directly test whether there is evidence of a network channel as this would entail observing an individual's or village's social network. However, we provide suggestive evidence that the network channel is trivial in our context, using two different measures of a village's network.

First, we use CDR data to define each village's network using call patterns between the cell tower servicing that village and other cell towers across Liberia. Broadly speaking, we consider two villages to be connected if there is significant call traffic between their

[35] Appendix Tables B9 and B10 present results that only include controls for topographic characteristics (elevation and slope). The results are qualitatively similar.

[36] Note in Appendix Table B9 that, although we find that the share of individuals reporting that they received EVD information and that a community taskforce visited the village are statistically significant, the overall information index remains statistically insignificant.

corresponding cell towers (Appendix Figure B3 shows for example that the core network for Ganta city likely lies in villages to the North and East of this tower). We then use Voronoi cells to draw catchment areas around each of the 152 unique Cellcom cell tower locations.[37] For each tower $k$, we then calculate the number of outgoing calls from that tower to all other towers across Liberia (including itself). We use several definitions of a network: (1) all villages within the catchment areas of tower $k$ plus the tower receiving the highest number of calls from $k$; (2) all villages within the catchment areas of tower $k$ plus the towers receiving the top five highest number of calls from $k$; (3) all villages within the catchment areas of tower $k$ plus the towers receiving 50% of all calls from $k$; (4) all villages within the catchment areas of tower $k$ plus the towers receiving 75% of all calls from $k$; and (5) all villages within the catchment areas of tower $k$ plus the towers receiving 90% of all calls from $k$.[38]

Second, we use historic Liberian clans as a way to define a village's closest social network. Officially, clans in Liberia are a third-tier administrative division.[39] However, clans also correspond to historical tribal chiefdoms that were merged into the state, throughout Liberian history, with chiefs simply assuming the role of agents of the central government (Nyei, 2014). Therefore, villages within the same clan are more likely to be socially interconnected than villages within another administrative unit. Based on the Census data (LISGIS, 2008), there are 631 clans in Liberia with an average of about 38 villages within each clan.

To assess whether the network channel is relevant, we estimate the following empirical model:

$$EVD_{ijt} = \alpha + \gamma EVD_{j,t-1} + \delta Match_{ij} \times EVD_{j,t-1} + \lambda_i + \nu_t + \epsilon_{ijt} \tag{4}$$

where $j$ indexes a network (CDR or clan-based definition). $EVD_{ijt}$ is an indicator for whether there is an EVD case in village $i$ of network $j$, in quarter $t$. $EVD_{j,t-1}$ is an indicator for

[37] Voronoi cells define the area for which any village within the cell is closest to that cell tower. Refer to Blumenstock et al. (2016) for earlier work using Voronoi cells to denote the catchment area of cell towers. Note that being within a tower's catchment area does not necessarily mean that the location gets coverage. It simply means that the location is closest to that specific tower.

[38] As an example, Appendix Figure B11 depicts the networks around the cell tower located in Ganta city, Nimba county, using these different definitions. Appendix Figure B12 also shows that most outgoing calls go to relatively few towers. Notice in panel (a) that for a given tower, 10 to 60% of all outgoing calls go to only one tower (the top called tower). However, only 0 to 5% of outgoing calls go to the 5th ranked tower. This indicates that the networks are generally small in terms of geographic extent.

[39] Administrative divisions in Liberia are county, district, clan, and village in this order.

whether there is at least one EVD case within village $i$'s network $j$ in the previous quarter. We define $Match_{ij}$ as a "coverage match" between village $i$ and any of the affected villages in the past quarter within village $i$'s network, i.e., $Match_{ij}$ equals 1 if village $i$ has cellphone coverage and at least one of the villages within $i$'s network $j$ having an EVD case in the past month also has coverage. $\lambda_i$ is a village fixed effect. $\nu_t$ is a quarter fixed effect. The village fixed effects $\lambda_i$ account for any time-invariant unobservables that may lead to endogenous selection of villages into EVD.[40]

In the presence of a network channel, we should expect the likelihood of an EVD case in village $i$ in quarter $t$ to decrease or increase if there is a "coverage match" between village $i$ and a village within $i$'s network with an EVD case at time $t-1$. Thus, we should expect coefficient $\delta$ to be nonzero if the network channel is non-trivial. We restrict the analysis to the pre-hotline/early-outbreak period when access to centralized relief efforts was limited and individuals likely relied on their network for help. This allows us to better isolate the effect of the network channel from other channels that likely relied on the existence of the hotline.

Table 8 presents the results. First, we find significant evidence of contagion within networks: Villages where another village within their network had an EVD case in the previous quarter are significantly more likely to report an EVD case in the current quarter. Although the evidence is stronger when using the CDR-defined networks (columns (1)-(5)), there is still some evidence of contagion within clans (column (6)). Not surprisingly, the magnitude of the contagion effect diminishes as we expand our definition of a network (0.0399 in column 1 *versus* 0.0310 in column 2, and moving from column 3 to column 5). This is expected if one considers that the likelihood of an interaction (and thus EVD transmission) is higher within a tighter definition of a network compared to a wider one. More importantly, we find no evidence of a network channel: for various definitions of a network, the coefficient estimates on the interaction term are small and statistically insignificant. This suggests that although there is evidence of a contagion effect, this effect does not significantly increase if two villages within the network happen to be connected.

[40] Given data sparsity in terms of EVD cases at the village-month level, we decide to aggregate cases to the village-quarter level. Note that we cannot estimate a coefficient for the $Match_{ij}$ variable by itself as it is collinear to the village fixed effect since coverage status does not change within our sample period.

### 6.4 Information Channel

We finally test whether access to cellphones lead individuals to have different access to (mis)information by the end of the epidemic. As part of our survey, we asked individuals whether they received daily information related to EVD (transmission, prevention, treatment). The survey also recorded each individual's beliefs on the origin of the outbreak. We use this latter information to classify individuals as *informed* if they reported that the epidemic originated at the border with Guinea and Sierra Leone or with traders from these areas. We classify individuals as *misinformed* if they reported that the origin of the outbreak was other institutions, distinguishing between government, foreign organizations/people (white people, UNMIL, foreign NGOs) and others (god, witches, the Fula, Mandingo, or Kissi) (see Gonzalez and Maffioli (2021) for more information on the classification).

We find almost no variation in the likelihood of receiving EVD-related information on a daily basis in our sample: more than 93% of respondents reports receiving information and this proportion does not change with coverage status.[41] Table 9 presents the results on the degree of knowledge about the origin of the epidemic. We do not find any evidence that individuals are more likely to be either informed or misinformed in cellphone coverage areas compared to non cellphone coverage areas.

The fact that we find no meaningful impacts of cellphone coverage on access to EVD-related information is not surprising in our context. First, access and use of radio in Liberia is much more ubiquitous than access to cellphone coverage (International Media Support, 2007). In our survey 80% of individuals report owning a radio, while in a national media survey 94% of urban dwellers and 91% of rural dwellers reported listening to the radio in the past week (Montez, 2010). Radio coverage is also widespread with several radio stations broadcasting at the national level and local community stations covering each county (International Media Support, 2007). In such a setting, the informational returns of cellphone access are likely trivial.

Second, recall that during the period of study the predominant cellphone technology available in Liberia was 2G (i.e., no data/Internet access through the phone). For instance, starting in 2014 only two towers in Monrovia and one tower near Roberts International

[41] We do not present any formal RD results on this given the lack of any meaningful variation in receiving information

Airport had 3G capabilities (i.e., slow Internet browsing ability) and none had 4G capabilities (fast Internet, TV and video streaming).[42] Thus, individuals at the time did not have access to video, social media, and other Internet content that can potentially be fertile ground for misinformation and EVD-related conspiracy theories. Therefore, it is not surprising that cellphone coverage within our context is not strongly associated with being misinformed.

## 7 Conclusion

Combining novel data on cellphone tower locations and Ebola cases in Liberia, we show that cellphone coverage contains the spread of the Ebola Virus Disease (EVD). Specifically, comparing villages at the margin of the signal strength threshold, we find that having access to cellphone coverage leads to a 10.8 percentage point reduction in the likelihood that a village has an EVD case. There is some indication that most of the effect is accrued to the introduction of a cellphone hotline designed to provide information about the disease and ease the provision of care. Using novel survey data collected after the epidemic, we assess the relative importance of several channels that may explain the observed relationship between cellphone coverage and epidemic containment. We provide evidence that the negative relationship between cellphone coverage and EVD is likely explained by facilitating access to treatment care, rather than improving access to prevention. This result is quite plausible as, in the short run, the returns to an additional ambulance or care center in the midst of a health crisis are likely higher and immediately realized. On the other hand, the effects of prevention may take longer to materialize as they essentially entail a change in health behavior.

Infectious disease outbreaks are still a major burden to low and middle-income countries (Holmes et al., 2017), and extreme events, such as health epidemics, are expected to remain a worldwide threat (United Nations Office for Disaster Risk Reduction, 2015). For instance, we are currently in the midst of a worldwide Coronavirus epidemic and a resurgence of Ebola in some African countries. Even though they might be unpredictable, the ultimate human and economic costs could be mitigated through appropriate governmental actions.

[42]These observations are made by the authors using the GSMA Collins Bartholomew Coverage Explorer GSMA (2014) which provides the type of coverage (2G,3G,4G) for GSM networks across the world.

Our findings show how something as simple and ubiquitous as a mobile phone can have positive externalities on economic development by allowing communities to better access health care treatment resources in times of crisis. From a policy perspective, it is fundamental for governmental stakeholders to know the relative effectiveness of potential tools, such as mobile phones, in mitigating the detrimental effect of infectious diseases. The results in this paper can help guide policymakers in choosing more efficient allocations of limited funds. For example, in normal times, longer term policies such as investments in the expansion of cellphone coverage to remote areas and information campaigns to prompt changes in preventive health behavior can be fruitful. During times of crisis, policymakers should take advantage of increased cellphone coverage to implement shorter-run, emergency policies such as hotlines, and measures that enhance access to treatment care and thus improve the effectiveness of the response. However, further research should explore specific interventions that directly take advantage of this technology and the effects of these tools in the context of other health epidemics.

## 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.

Agarwal, S., H. B. Perry, L.-A. Long, and A. B. Labrique (2015). Evidence on feasibility and effective use of mhealth strategies by frontline health workers in developing countries: systematic review. *Tropical Medicine and International Health 20(8)*, 1003–1014.

Aker, J. (2010). Information from Markets Near and Far: Mobile Phones and Agricultural Markets in Niger. *American Economic Journal: Applied Economics*, 46–59.

Aker, J. and M. Fafchamps (2014). Mobile phone coverage and producer markets: evidence from West Africa. *World Bank Economic Review 29*(2), 262 – 292.

Aker, J. and C. Ksoll (2018). Can ABC Lead to Sustained 123? The Medium-Term Effects of a Technology-Enhanced Adult Education Program. *Economic Development and Cultural Change*.

Aker, J., C. Ksoll, and T. J. Lybbert (2012). Can mobile phones improve learning? Evidence from a field experiment in Niger. *American Economic Journal: Applied Economics 4*(4), 94–120.

Aker, J. and I. M. Mbiti (2010). Mobile phones and economic development in Africa. *Journal of Economic Perspectives 24*(3), 207–232.

Alexander, K. et al. (2015). What factors might have led to the emergence of ebola in west africa? *PLoS Neglected Tropical Diseases 9*(6).

Allgaier, J. and A. L. Svalastog (2015). The communication aspects of the ebola virus disease outbreak in western africa–do we need to counter one, two, or many epidemics? *Croatian Medical Journal 56*(5), 496.

Armand, A., P. Atwell, and J. F. Gomes (2020). The reach of radio: Ending civil conflict through rebel demobilization. *American economic review 110*(5), 1395–1429.

Baker, A. (2014). People are prank calling liberia's ebola hotline. *Time*.

Bengtsson, L., J. Gaudart, X. Lu, S. Moore, E. Wetter, K. Sallah, S. Rebaudet, and R. Piarroux (2015). Using mobile phone data to predict the spatial spread of cholera. *Scientific reports 5*, 8923.

Blumenstock, J. E. (2012). Inferring patterns of internal migration from mobile phone call records: evidence from rwanda. *Information Technology for Development 18*(2), 107–125.

Blumenstock, J. E., N. Eagle, and M. Fafchamps (2016). Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters. *Journal of Development Economics 120*, 157–181.

Blumenstock, J. E., M. Fafchamps, and N. Eagle (2011). Risk and reciprocity over the mobile phone network: evidence from rwanda. *Available at SSRN 1958042*.

Braun, R., C. Catalani, J. Wimbush, and D. Israelski (2013). Community health workers and mobile technology: A systematic review of the literature. *PLoS One 8(6)e65772*.

Burbank, J. L., J. Andrusenko, J. S. Everett, and W. T. Kasch (2013). *Wireless networking: Understanding internetworking challenges*. John Wiley & Sons.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Card, D., A. Mas, and J. Rothstein (2008). Tipping and the dynamics of segregation. *The Quarterly Journal of Economics 123*(1), 177–218.

Carey, J. M., V. Chi, D. Flynn, B. Nyhan, and T. Zeitzoff (2020). The effects of corrective information about disease epidemics and outbreaks: Evidence from zika and yellow fever in brazil. *Science advances 6*(5), eaaw7449.

Cattaneo, M. D., M. Jansson, and X. Ma (2018). Manipulation testing based on density discontinuity. *The Stata Journal 18*(1), 234–261.

Cattaneo, M. D., M. Jansson, and X. Ma (2019). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 1–7.

Crabtree, C. and H. L. Kern (2018). Using electromagnetic signal propagation models for radio and television broadcasts: An introduction. *Political Analysis 26*(3), 348–355.

D'Ambrosio, M. V., M. Bakalar, S. Bennuru, C. Reber, A. Skandarajah, L. Nilsson, N. Switz, J. Kamgno, S. Pion, M. Boussinesq, T. Nutman, and D. A. Fletcher (2015). Point-of-care quantification of blood-borne filarial parasites with a mobile phone microscope. *Science translational medicine 7*, 286re4.

Eppink, D. and W. Kuebler (1994). Tirem/sem handbook.

Fallah, M. P., L. A. Skrip, S. Gertler, D. Yamin, and A. P. Galvani (2015). Quantifying poverty as a driver of ebola transmission. *PLoS Neglected Tropical Diseases 9*(12).

Farahani, S. (2008). Chapter 4: Transceivers requirements. *ZigBee Wireless Networks and Transceivers*.

Feng, S., K. A. Grepin, and R. Chunara (2018). Tracking health seeking behavior during an Ebola outbreak via mobile phones and SMS. *NPJ Digit Med 1*(51).

Freifeld, C. C., R. Chunara, S. R. Mekaru, E. H. Chan, T. Kass-Hout, A. A. Iacucci, and J. S. Brownstein (2010). Participatory epidemiology: use of mobile phones for community-based health reporting. *PLoS Medicine 7 (12)*.

Gagliarducci, S., M. G. Onorato, F. Sobbrio, and G. Tabellini (2020). War of the waves: Radio and resistance during world war ii. *American Economic Journal: Applied Economics 12*(4), 1–38.

Gonzalez, R. and E. M. Maffioli (2021). Profile of a Conspiracy Theorist: The Role of Government Trust and Technology on Misinformation during an Epidemic. *Available at SSRN 3688576*.

Gonzalez, R. M. (2021). Cell phone access and election fraud: Evidence from a spatial regression discontinuity design in afghanistan. *American Economic Journal: Applied Economics 13*(2), 1–51.

GSMA (2014). Mobile coverage explorer. *Collins Bartholomew* (Available at: https://www.collinsbartholomew.com/mobile-coverage-maps/mobile-coverage-explorer).

GSMA (2019). Operator Acceptance Values for Device Antenna Performance Version 4.0. Technical report.

Guriev, S., N. Melnikov, and E. Zhuravskaya (2020). 3g internet and confidence in government. *The Quarterly Journal of Economics*.

Hampton, K. N., L. F. Sessions, and E. J. Her (2011). Core networks, social isolation, and new media: How internet and mobile phone use is related to network size and diversity. *Information, Communication and Society 14*(1), 130–155.

Holmes, K. K., S. Bertozzi, B. R. Bloom, P. Jha, H. Gelband, L. M. DeMaria, and S. Horton (2017). *Major Infectious Diseases. 3rd edition.* Washington (DC): The International Bank for Reconstruction and Development and The World Bank.

Hymowitz, D. (2017). State of Emergency: Lessons from how government fought Ebola. Technical report, Tony Blair Institute for Global Change.

IFRC (2014). Emergency appeal operation update ebola virus disease emergency appeals (liberia, sierra leone, guinea, nigeria, senegal and africa coordination). *International Federation of Red Cross and Red Crescent*.

International Media Support (2007). A review of media support in the post-conflict transitional period and recommendations for future actions: Strengthening liberia's media. *The partnership for media and conflict prevention in West Africa*.

Jack, W. and T. Suri (2011). Mobile money: The economics of M-PESA.

Jack, W. and T. Suri (2014). Risk sharing and transactions costs: evidence from Kenya's mobile money revolution. *American Economic Review 104*(1), 183–223.

Jack, W., T. Suri, and R. Townsend (2010). Monetary theory and electronic money: Reflections on the kenyan experience. *FRB Richmond Economic Quarterly*.

JAXA (2016). Alos global digital surface model alos world 3d 30m (aw3d30), japan aerospace exploration agency.

Jensen, R. (2007). The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector. *Quarterly Journal of Economics 122*, 879–924.

Kirsch, T. D., H. Moseson, M. Massaquoi, T. G. Nyenswah, R. Goodermote, I. Rodriguez-Barraquer, J. Lessler, D. A. T. Cumings, and D. H. Peters (2017). Impact of interventions and the incidence of ebola virus disease in liberia - implications for future epidemics. *Health Policy and Planning 32*, 205–214.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Krishna, A. and T. L. Thompson (2019). Misinformation about health: a review of health communication and misinformation scholarship. *American Behavioral Scientist*, 0002764219878223.

Lazaridis, P., A. Bizopoulos, S. Kasampalis, J. Cosmas, and Z. D. Zaharis (2013). Evaluation of prediction accuracy for the longley-rice model in the fm and tv bands.

Liberian Ministry of Health (2017). 2014 annual report.

LISGIS (2008). 2008 national population and housing census. *Liberia Institute of Statistics and Geo-Information Services (LISGIS)*. Available at: https://www.lisgis.net/page_info.php?7d5f44532cbfc489b8db9e12e44eb820=MzQy.

Longley, A. G. and P. L. Rice (1968). Prediction of tropospheric radio transmission loss over irregular terrain. a computer method-1968.

LTA (2014). Annual report 2013-2014. *Liberia Telecommunications Authority*.

Lu, X., L. Bengtssona, and P. Holmea (2012). Predictability of population displacement after the 2010 Haiti earthquake. *PNAS 109*(29), 11576–11581.

Maffioli, E. M. (2020). Collecting data during an epidemic: A novel mobile phone research method. *Journal of International Development*.

Maffioli, E. M. (2021). The Political Economy of Health Epidemics: Evidence from the Ebola Outbreak. *Journal of Development Economics 151*.

Manacorda, M. and A. Tesei (2020). Liberation technology: Mobile phones and political mobilization in africa. *Econometrica 88*(2), 533–567.

Montez, D. (2010). Community radio a vital resource for liberians. *AudienceScapes: Intermedia Knowledge Center*.

Nyei, I. (2014). Decentralizing the state in liberia: The issues, progress and challenges. *Stability: International Journal of Security and Development 3*(1).

Nyenswah, T. G., F. Kateh, L. Bawo, M. Massaquoi, M. Gbanyan, M. Fallah, et al. (2016). Ebola and Its Control in Liberia, 2014–2015. *Emerging Infectious Diseases 22*(2).

Obasola, O. I., I. Mabawonku, and I. Lagunju (2015). A review of e-health interventions for maternal and child health in sub-sahara africa. *Matern Child Health Journal 19*, 1813–1824.

Olken, B. A. (2009). Do television and radio destroy social capital? evidence from indonesian villages. *American Economic Journal: Applied Economics 1*(4), 1–33.

Onyeonoro, U. U., U. C. Ekpemiro, C. Abali, and H. I. Nwokeukwu (2015). Ebola epidemic-the nigerian experience. *The Pan African Medical Journal 22*(Suppl 1).

Open Topography (2017). Alos world 3d - 30m.

Ortiz-Martínez, Y. and L. F. Jiménez-Arcia (2017). Yellow fever outbreaks and twitter: Rumors and misinformation. *American Journal of Infection Control 45*(7), 816–817.

Oyeyemi, S. O., E. Gabarron, and R. Wynn (2014). Ebola, twitter, and misinformation: a dangerous combination? *Bmj 349*, g6178.

Pathak, R., D. R. Poudel, P. Karmacharya, A. Pathak, M. R. Aryal, M. Mahmood, and A. A. Donato (2015). Youtube as a source of information on ebola virus disease. *North American journal of medical sciences 7*(7), 306.

Pew Research Center (2011). Social networking sites and our lives: How people's trust, personal relationships, and civic and political involvement are connected to their use of social networking sites and other technologies. Technical report.

Pew Research Center (2019). In emerging economies, smartphone and social media users have broader social networks. Technical report.

Porter, J. and P. Yu (2015). Regression discontinuity designs with unknown discontinuity points: Testing and estimation. *Journal of Econometrics 189*(1), 132–147.

Qiu, P. (2011). Jump regression analysis. *International Encyclopedia of Statistical Science*.

Razally, F. (2015). Mobile Handset Testing: A Report for OFCOM, The UK Communication Regulator. Technical report.

Roberts, H., B. Seymour, S. A. Fish, E. Robinson, and E. Zuckerman (2017). Digital health communication and global public influence: a study of the ebola epidemic. *Journal of Health Communication 22*(sup1), 51–58.

Ruble, K. (2015). The village that beat ebola: How one liberian community avoided the outbreak. *Vice News*.

Sacks, J. A., E. Zehe, C. Redick, A. Bah, K. Cowger, M. Camara, A. Diallo, A. Nasser, I. Gigo, R. S. Dhillon, and A. Liua (2015). Introduction of mobile health tools to support ebola surveillance and contact tracing in guinea. *Global Health: Science and Practice 2 (4)*, 646–659.

Seybold, J. S. (2005). *Introduction to RF propagation*. John Wiley & Sons.

Shapiro, J. N. and N. B. Weidmann (2015). Is the phone mightier than the sword? cellphones and insurgent violence in iraq. *International Organization*, 247–274.

Spokoiny, V. G. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *The annals of statistics 26*(4), 1356–1378.

United Nations Office for Disaster Risk Reduction (2015). The human cost of weather-related disasters: 1995-2015. Technical report.

Venkatraman, A., D. Mukhija, N. Kumar, and S. Nagpal (2016). Zika virus misinformation on the internet. *Travel medicine and infectious disease 14*(4), 421–422.

Weber, S. and M. Péclat (2017). A simple command to calculate travel distance and travel time. *The Stata Journal 17*(4), 962–971.

Wesolowski, A., T. Qureshi, M. F. Boni, P. R. Sundsøy, M. A. Johansson, S. B. Rasheed, K. Engø-Monsen, and C. O. Buckee (2015). Impact of human mobility on the emergence of dengue epidemics in pakistan. *Proceedings of the National Academy of Sciences 112*(38), 11887–11892.

White House (2014). White House, Office of the Press Secretary. 2014b. White House Update on U.S. Response to Ebola, 02 Dec. 2014. Available from: https://www.whitehouse.gov/the-press-office/2014/12/02/fact-sheet-update-ebola-response. Technical report.

World Bank (2014). Press release. World Bank Group to Nearly Double Funding in Ebola Crisis to $400 Million, September 25 2014. Available from: http://www.worldbank.org/en/news/press-release/ 2014/09/25/world-bank-group-nearly-double-funding-ebola-crisis-400-million. Technical report.

World Health Organization (2014). Ebola: Experimental therapies and rumoured remedies.

World Health Organization (2016). Situation report, liberia: June 2016. Technical report.

Yang, C., J. Yang, X. Luo, and P. Gong (2009). Use of mobile phones in an emergency reporting system for infectious disease surveillance after the sichuan earthquake in china. *Bulletin of the World Health Organization 87*, 619–632.

# Figures and Tables

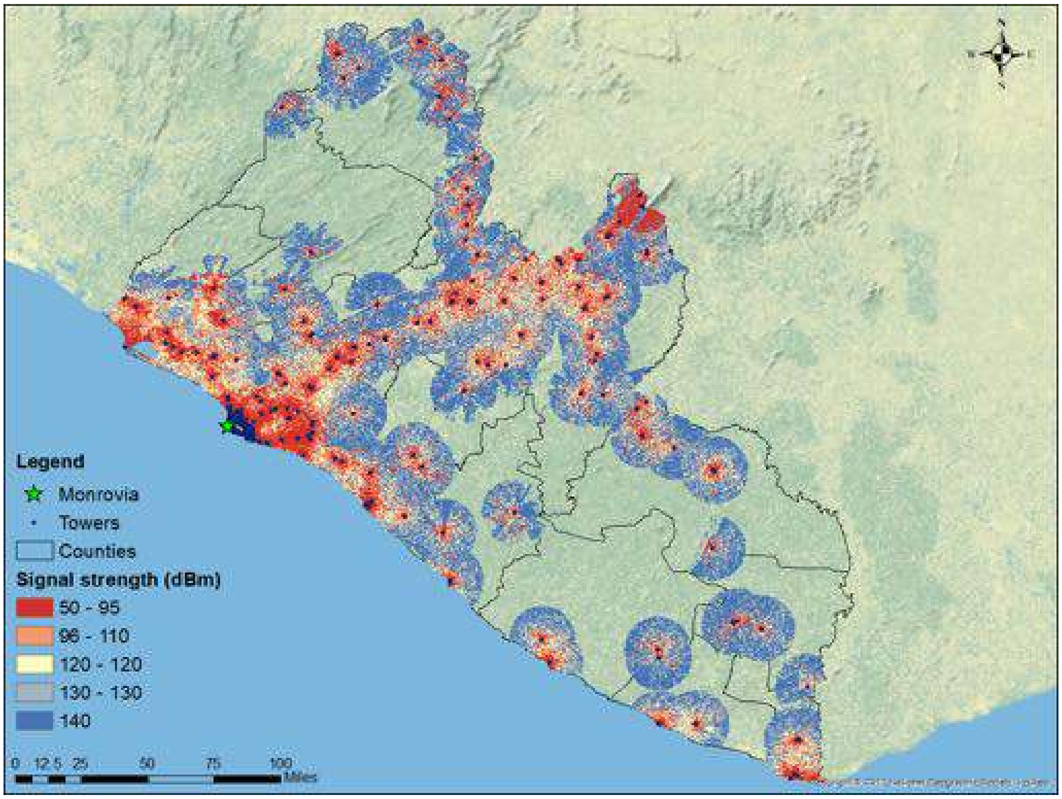
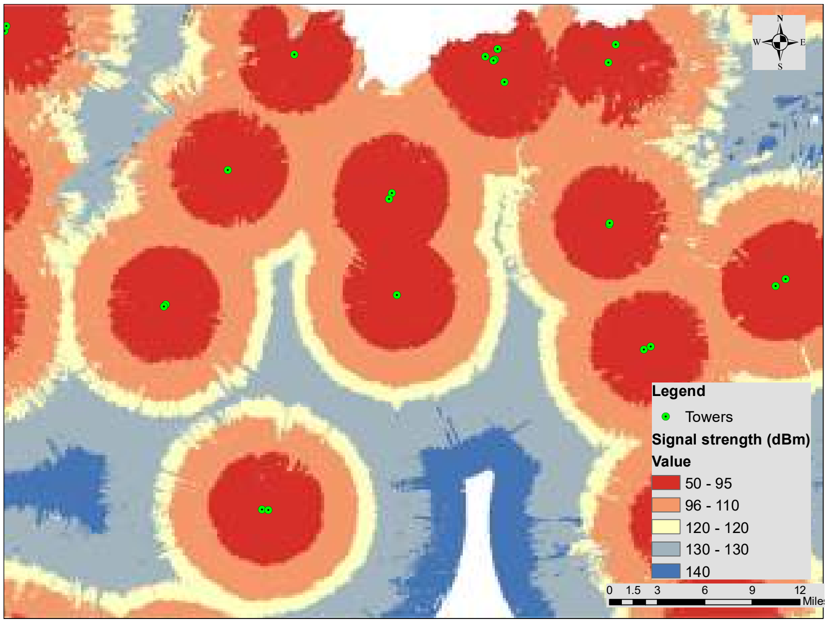
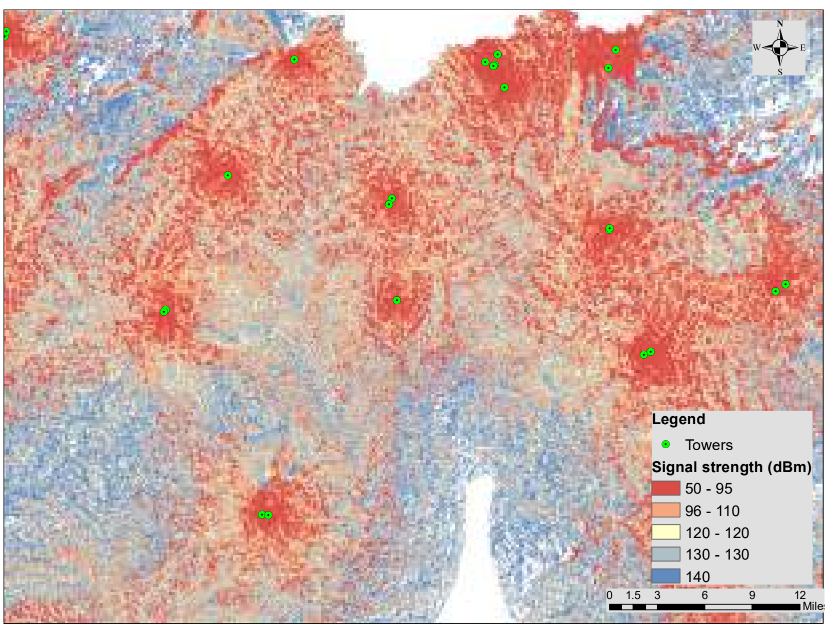



Figure 1: Irregular Terrain Model, Liberia (2013)

*Notes:* Cellphone towers' location obtained from the Liberia Telecommunications Authority (LTA). Estimates of the ITM model described in Section 3.2. Lower dBm values mean higher signal strength (i.e., more coverage).



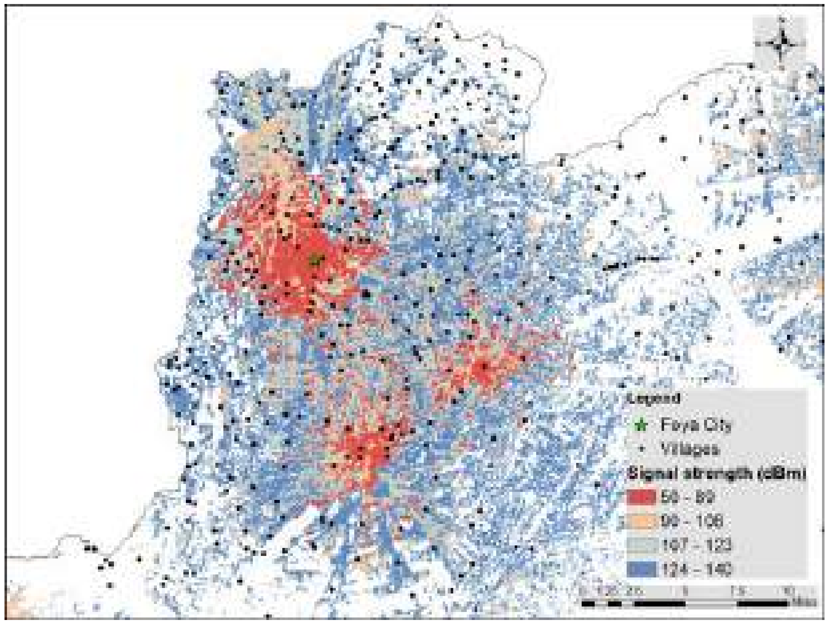

(a) “Free-space” ITM



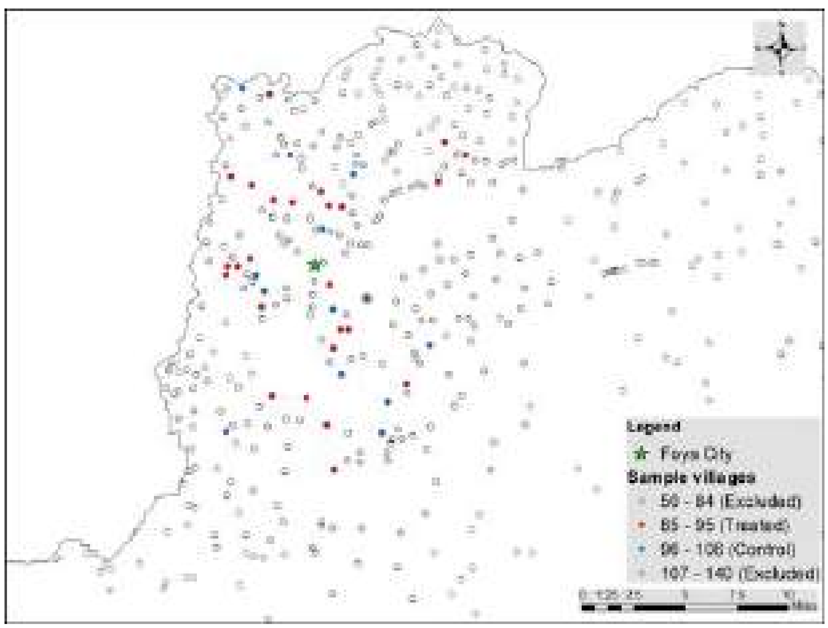

(b) ITM with Topography

Figure 2: ITM “Free-space” versus ITM with Topography, Liberia (2013)

*Notes*: Green dots give the location of towers. Sample along the border of Nimba and Bong counties in Northern Liberia.



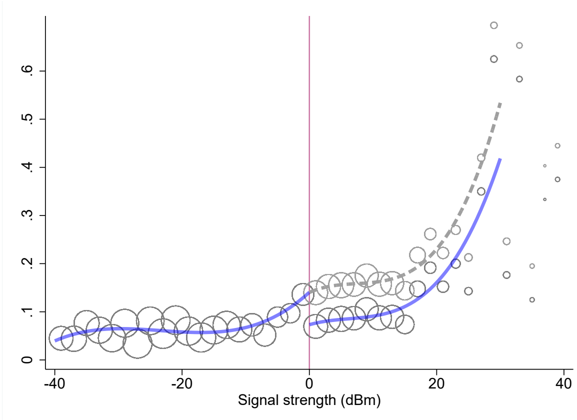

(a) ITM detailed



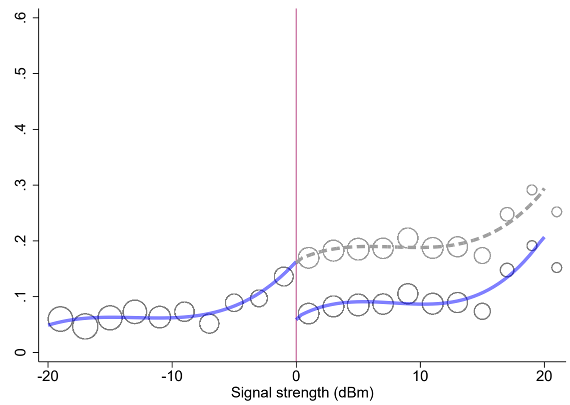

(b) Villages within 10dBm Bandwidth of Cutoff

Figure 3: Irregular Terrain Model with Village Sample

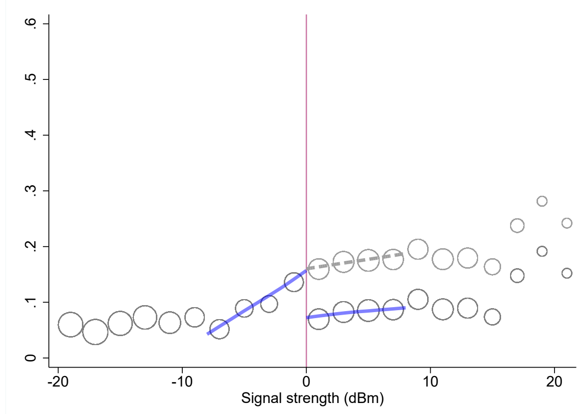
*Notes:* Cellphone towers' location obtained from the Liberia Telecommunications Authority (LTA). Estimates of the ITM model described in Section 3.2. Lower dBm values mean higher signal strength (i.e., more coverage). Dots indicate the location of villages.




(a) Wide bandwidth (40dBm)




(b) Narrow bandwith (20dBm)




(c) Local Linear Regression

Figure 4: Regression Discontinuity (RD) Plots for Likelihood of EVD Case

*Notes:* Dots give EVD likelihood for each bin of signal strength (dBm). Dot size weighted by the number of observations within each bin. "Signal strength (dBm)" refers to the received power or signal strength, measured in decibel milliwatts, at the village level. Signal strength is normalized so that negative values mean no coverage (i.e., not enough signal strength). Bin width is 2 dBm. Window of analysis is -40 to 40dBm (panel a) and -20 to 20 dBm (panels b and c). Solid lines give the predicted values from a regression of the outcome variable on a third degree polynomial in distance to threshold that uses a triangular kernel (panels a and b) and a linear trend using a triangular kernel (panel c). Light gray dots and dashed lines give a representation of the potential outcomes by shifting upwards the graph for positive dBm values to the point where the trends on each side of zero intersect.

Table 1: Determinants of Coverage

| | Dep. variable: Signal strength (dBm) | | |
|---|---|---|---|
| | Full sample | Within 20 dBm | Within 10 dBm |
| | (1) | (2) | (3) |
| *Topographic controls* | | | |
| Elevation (m) | -0.082*** | -0.032** | 0.005 |
| | (0.024) | (0.014) | (0.009) |
| Slope (%) | -0.283 | 1.129*** | 0.828*** |
| | (0.455) | (0.379) | (0.256) |
| *Demographic controls* | | | |
| Household size | -0.194 | 0.014 | 0.118 |
| | (0.171) | (0.131) | (0.087) |
| Population (log) | 0.880*** | 0.225 | -0.013 |
| | (0.166) | (0.162) | (0.117) |
| Female (%) | -0.367 | 1.805 | 1.474 |
| | (2.456) | (3.029) | (1.775) |
| Married (%) | -4.267* | -1.750 | 1.236 |
| | (2.170) | (1.679) | (1.481) |
| Christian (%) | 1.974 | 3.553** | -0.340 |
| | (2.593) | (1.543) | (1.603) |
| Muslim (%) | 1.655 | -0.263 | -1.721 |
| | (3.180) | (2.473) | (1.982) |
| African religion (%) | 6.102 | 6.790 | -0.579 |
| | (5.602) | (6.518) | (7.128) |
| Kpelle (%) | -0.040 | -1.410 | -1.190 |
| | (1.159) | (1.430) | (0.946) |
| Bassa (%) | 3.648 | 1.031 | 0.497 |
| | (2.229) | (1.771) | (1.163) |
| *Economic controls* | | | |
| Primary education (%) | -0.347 | 0.142 | -1.199 |
| | (1.665) | (1.269) | (1.200) |
| Secondary education (%) | 5.027** | 2.097* | 0.925 |
| | (2.091) | (1.107) | (1.137) |
| Owns house (%) | 1.084 | -0.498 | -0.403 |
| | (1.189) | (0.858) | (0.568) |
| House condition: Good (%) | 9.476*** | 4.908*** | 0.858 |
| | (2.175) | (1.332) | (0.941) |
| Assets ownership (%) | 2.350 | 1.413 | 0.505 |
| | (3.844) | (2.828) | (1.621) |
| Distance to Monrovia (km) | -0.098** | -0.013 | -0.020 |
| | (0.047) | (0.049) | (0.024) |
| Distance to closest city (km) | -0.306*** | -0.064 | -0.035 |
| | (0.113) | (0.088) | (0.049) |
| Observations | 7,014 | 3,839 | 1,913 |
| P-value for F-test for joint significance: | | | |
| Topographic controls | 0.00 | 0.00 | 0.00 |
| Demographic controls | 0.00 | 0.01 | 0.35 |
| Economic controls | 0.00 | 0.00 | 0.79 |
| Demographic and economic controls | 0.00 | 0.00 | 0.15 |

*Notes:* Standard errors clustered at district level. All specification include district fixed effects. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table 2: Effect of Coverage on Likelihood of EVD Case

| | Dep. Variable = $\mathbb{1}\{$Number of Reported EVD cases $> 0\}$ | | | | | |
|---|---|---|---|---|---|---|
| | Baseline | Controls | Topography Poly | Polynomial RD | Kernel choice | Probit model |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Coverage | -0.108** | -0.100** | -0.110** | -0.096*** | -0.106*** | -0.429** |
| | (0.048) | (0.042) | (0.048) | (0.035) | (0.041) | (0.206) |
| Mean outside coverage | 0.09 | 0.09 | 0.09 | 0.06 | 0.09 | 0.06 |
| Bandwidth (dBm) | 8.99 | 8.17 | 9.07 | 50.00 | 8.21 | 8.99 |
| Observations | 1547 | 1547 | 1741 | 7014 | 1547 | 1547 |
| Districts | 83 | 83 | 84 | 115 | 83 | 83 |
| Marginal Effect | - | - | - | - | - | -0.070 |

*Notes:* Columns (1) and (2) present estimates of $\beta$ using a local linear regression specification of Equation (1). All specification include controls for elevation and slope. Column (2) adds controls for average household size, population, female, married, Christian, Muslim, African religion, Kpelle, Bassa, primary education, literacy, house ownership and quality, an index of asset ownership, distance to Monrovia, and distance to closest main city. Optimal bandwidth chosen as in Calonico et al. (2014). Column (3) uses topography polynomial: $h(\mathbf{G}_i) = \rho_1 elev_i + \rho_2 elev_i^2 + \rho_3 slope_i + \rho_4 elev_i \times slope_i + \rho_5 elev_i^2 \times slope_i + \rho_6 slope_i^2 + \rho_7 elev_i \times slope_i^2 + \rho_8 elev_i^2 \times slope_i^2$. It also includes distance to tower as additional control. Column (4) uses a wider bandwidth of 50 dBm and a third degree polynomial specification for $f(\tilde{R}_i)$. Column (5) uses a rectangular kernel. Column (6) estimates a probit specification of Equation (1). Optimal bandwidths chosen as in Calonico et al. (2014) except for column (4) which uses a fixed, wider bandwidth. Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table 3: Effect of Coverage on Likelihood of EVD, by Past EVD Exposure

| | Dep. Variable = $\mathbb{1}\{\text{Number of Reported EVD cases} > 0\}$ | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Coverage$_i$ | -0.0069** | -0.0059** | | 0.0005 | 0.0013 |
| | (0.003) | (0.002) | | (0.002) | (0.002) |
| EVD$_{j(i),t-1}$ | | | 0.0056** | 0.0167** | 0.0159** |
| | | | (0.002) | (0.007) | (0.007) |
| Coverage$_i$ $\times$ EVD$_{j(i),t-1}$ | | | | -0.0191*** | -0.0191*** |
| | | | | (0.006) | (0.006) |
| Mean outside coverage | 0.007 | 0.007 | 0.006 | 0.007 | 0.007 |
| Observations | 33079 | 33079 | 31338 | 31338 | 31338 |
| Bandwidth (dBm) | 9.00 | 9.00 | 9.00 | 9.00 | 9.00 |
| Districts | 84 | 84 | 84 | 84 | 84 |
| District FE | Yes | Yes | Yes | Yes | Yes |
| Controls | No | Yes | Yes | No | Yes |

*Notes:* Coverage$_i$ equal 1 if village $i$ has coverage. EVD$_{j(i),t-1}$ equals 1 if there is one or more EVD cases in at least one village within village $i$'s district in the past month. Standard errors clustered at district level. Columns (2), (3), and (5) include controls for average household size, population, female, married, Christian, Muslim, African religion, Kpelle, Bassa, primary education, literacy, house ownership and quality, an index of asset ownership, distance to Monrovia, and distance to closest main city. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table 4: Controlling for “Free-Space” Signal Strength and Falsification Tests

| | Dep. Variable = $\mathbb{1}\{$Number of Reported EVD cases $> 0\}$ | | | | | |
|---|---|---|---|---|---|---|
| | Controlling for Free-space signal strength | | | | Falsification tests | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Coverage | -0.087** | -0.084* | -0.077** | -0.081** | | |
| | (0.038) | (0.041) | (0.037) | (0.040) | | |
| Free-space coverage | -0.013 | | -0.044* | -0.045* | -0.003 | |
| | (0.020) | | (0.024) | (0.026) | (0.040) | |
| Free-space signal strength | | 0.000 | -0.000 | -0.000 | 0.002 | |
| | | (0.001) | (0.001) | (0.001) | (0.006) | |
| Free-space coverage × Free-space signal strength | | | 0.003 | 0.003 | -0.008 | |
| | | | (0.002) | (0.002) | (0.007) | |
| Coverage (2015) | | | | | | 0.084 |
| | | | | | | (0.112) |
| Signal strength (2015) | | | | | | 0.001 |
| | | | | | | (0.002) |
| Mean outside coverage | 0.093 | 0.101 | 0.094 | 0.093 | 0.094 | 0.057 |
| Bandwidth (dBm) | 8.35 | 7.99 | 9.14 | 8.93 | 9.36 | - |
| Observations | 1547 | 1352 | 1720 | 1528 | 1540 | 383 |
| Districts | 83 | 81 | 84 | 83 | 84 | 53 |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Topography Polynomial | No | No | No | Yes | No | No |

*Notes:* Columns (1)-(5) present estimates using a local linear regression specification of Equation (1). Column (1) adds controls for coverage under free-space. Column (2) adds controls for free-space signal strength. Columns (3) and (4) add the full interaction of free-space coverage and strength. Column (5) estimates the local linear regression specification of Equation (1) using free-space coverage. Column (6) estimates the effect of coverage in 2015 excluding districts that had coverage in 2013, including district fixed-effects. All specifications include controls for elevation, slope, average household size, population, female, married, Christian, Muslim, African religion, Kpelle, Bassa, primary education, literacy, house ownership and quality, an index of asset ownership, distance to Monrovia, and distance to closest main city. Optimal bandwidth in Columns (1)-(5) chosen as in Calonico et al. (2014). Column (4) adds topography polynomial: $h(\mathbf{G}_i) = \rho_1 elev_i + \rho_2 elev_i^2 + \rho_3 slope_i + \rho_4 elev_i \times slope_i + \rho_5 elev_i^2 \times slope_i + \rho_6 slope_i^2 + \rho_7 elev_i \times slope_i^2 + \rho_8 elev_i^2 \times slope_i^2$. It also includes distance to tower as additional control. Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table 5: Effect of Coverage on Likelihood of EVD: Before and After Hotline Introduction

| | Dep. Variable = $\mathbb{1}\{$Number of Reported EVD cases $> 0\}$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pre-Hotline | | | | | | Post-Hotline | | | | | |
| | Baseline | Controls | Topography Poly | Poly RD | Kernel choice | Probit model | Baseline | Controls | Topography Poly | Poly RD | Kernel choice | Probit model |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Coverage | -0.01 | -0.00 | -0.01 | -0.00 | -0.01 | -0.01 | -0.11** | -0.10*** | -0.11** | -0.10*** | -0.09*** | -0.44*** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.44) | (0.04) | (0.04) | (0.04) | (0.03) | (0.03) | (0.15) |
| Mean outside coverage | 0.010 | 0.012 | 0.012 | 0.012 | 0.010 | 0.010 | 0.087 | 0.088 | 0.087 | 0.056 | 0.087 | 0.054 |
| Bandwidth (dBm) | 10.33 | 11.23 | 11.21 | 50.00 | 8.81 | . | 9.00 | 8.31 | 9.14 | 50.00 | 9.40 | . |
| Observations | 1913 | 2139 | 2139 | 7014 | 1547 | 1547 | 1741 | 1547 | 1741 | 7014 | 1741 | 1741 |
| Districts | 86 | 87 | 87 | 115 | 83 | 83 | 84 | 83 | 84 | 115 | 84 | 84 |
| Marginal Effect | | | | | | -0.000 | | | | | | -0.069 |

*Notes:* Columns (1), (2), (7), and (8) present estimates of $\beta$ using a local linear regression specification of Equation (1). All specification include controls for elevation and slope. Columns (2) and (8) add controls for average household size, population, female, married, Christian, Muslim, African religion, Kpelle, Bassa, primary education, literacy, house ownership and quality, an index of asset ownership, distance to Monrovia, and distance to closest main city. Columns (3) and (9) use topography polynomial: $h(\mathbf{G}_i) = \rho_1 elev_i + \rho_2 elev_i^2 + \rho_3 slope_i + \rho_4 elev_i \times slope_i + \rho_5 elev_i^2 \times slope_i + \rho_6 slope_i^2 + \rho_7 elev_i \times slope_i^2 + \rho_8 elev_i^2 \times slope_i^2$. It also includes distance to tower as additional control. Columns (4) and (10) use a wider bandwidth of 50 dBm and a third degree polynomial specification for $f(\tilde{R}_i)$. Columns (5) and (11) use a rectangular kernel. Columns (6) and (12) use a probit specification of Equation (1). Optimal bandwidths chosen as in Calonico et al. (2014) except for columns (4) and (10) which use a fixed, wider bandwidth. Standard errors clustered at district level in all specifications. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table 6: Effect of Coverage on Preventive Care

| | Contact tracing | Explain EVD | Hygiene meetings | Bring info (Task-force) | Explain burials | Prevention material | Preventive care index |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Coverage | 0.102 | 0.123 | -0.173 | 0.135 | 0.090 | 0.079 | 0.049 |
| | (0.084) | (0.086) | (0.110) | (0.088) | (0.092) | (0.111) | (0.131) |
| Mean outside coverage | 0.200 | 0.897 | 0.465 | 0.282 | 0.174 | 0.779 | 0.047 |
| Bandwidth (dBm) | 8.80 | 14.89 | 17.12 | 8.94 | 11.77 | 16.46 | 15.17 |
| Observations | 268 | 434 | 551 | 296 | 328 | 518 | 447 |
| Villages | 86 | 166 | 203 | 86 | 122 | 189 | 175 |

*Notes:* Results present estimates of $\beta$ using a local linear regression specification of Equation (1) on preventive care variables. The observations are respondents in the survey sample (2016). Optimal bandwidth chosen as in Calonico et al. (2014). The information index is constructed following Kling et al. (2007). All specifications include controls for elevation, slope, sex, age, urban, secondary education level, and categories for Kpelle, Bassa, and other tribes. Standard errors clustered at the village level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table 7: Effect of Coverage on Treatment Care

| | Take sick | Take dead | Ambulance on-time | CCCs within 10km | Treatment care index |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Coverage | 0.223*** | -0.006 | 0.206*** | 0.416* | 0.340** |
| | (0.036) | (0.044) | (0.076) | (0.235) | (0.146) |
| Mean outside coverage | 0.123 | 0.070 | 0.835 | 0.455 | -0.036 |
| Bandwidth (dBm) | 8.13 | 11.17 | 12.35 | 9.40 | 8.00 |
| Observations | 268 | 328 | 385 | 317 | 268 |
| Villages | 86 | 122 | 139 | 100 | 86 |

*Notes:* Results present estimates of $\beta$ using a local linear regression specification of Equation (1) on treatment care variables. The observations are respondents in the survey sample (2016). Optimal bandwidth chosen as in Calonico et al. (2014). The care index is constructed following Kling et al. (2007). All specifications include controls for elevation, slope, sex, age, urban, and education level, and categories for Kpelle, Bassa, and other tribes. Standard errors clustered at the village level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table 8: Likelihood of EVD within the Network

| | Dep. Variable = $\mathbb{1}\{\text{Number of Reported EVD cases} > 0\}$ | | | | | |
|---|---|---|---|---|---|---|
| | Network: Call Detail Records (CDR) | | | | | |
| | Top tower (1) | Top 5 towers (2) | 50% call share (3) | 75% call share (4) | 90% call share (5) | Network: Clan (6) |
| $\text{EVD}_{j,t-1}$ | 0.0399*** | 0.0310*** | 0.0406*** | 0.0179** | 0.0095** | 0.0338 |
| | (0.007) | (0.009) | (0.007) | (0.008) | (0.004) | (0.028) |
| $\text{Match}_{ij} \times \text{EVD}_{j,t-1}$ | 0.0061 | 0.0064 | 0.0109 | 0.0112 | 0.0085 | -0.0030 |
| | (0.033) | (0.024) | (0.036) | (0.016) | (0.006) | (0.045) |
| Observations | 18864 | 18864 | 18864 | 18864 | 18864 | 19372 |
| Villages | 9432 | 9432 | 9432 | 9432 | 9432 | 9686 |
| Clusters | 96 | 96 | 96 | 96 | 96 | 631 |
| Village FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Quarter FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | No | No | No | No | No | No |

*Notes:* Dependent variable is a dummy indicator for whether there is at least one EVD case in village $i$ in quarter $t$. $\text{EVD}_{j,t-1}$ equals 1 if there is at least one EVD case within village $i$'s network $j(i)$ in the past quarter. $\text{Match}_{ij}$ equals 1 if there is a cellphone coverage match between village $i$ and any of the affected villages in the past quarter within village $i$'s network (i.e., village $i$ has coverage and at least one EVD-affected village within $i$'s network also has coverage). Columns 1-5 define village $i$'s network using call detail records (CDR) data between village $i$'s cell tower and other cell towers across Liberia. Columns 1 and 2 define $i$'s network as all villages within the service area of the top and top-5 towers receiving calls originating from $i$'s tower. Columns 3-5 define $i$'s network as all villages within the service area of cell towers receiving 50%, 75%, and 90%, respectively, of all calls originating from $i$'s tower. CDR data obtained from Cellcom Liberia. Column 6 defines $i$'s network as all villages within the same clan as village $i$. Standard errors clustered at the network level. All specifications include village and quarter fixed effects. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table 9: Effect of Coverage on (Mis)information

| | Informed (1) | Misinformed (2) | Gov (3) | Foreign (4) | Other (5) |
|---|---|---|---|---|---|
| Coverage | 0.102 | -0.097 | -0.027 | 0.011 | -0.088 |
| | (0.124) | (0.107) | (0.089) | (0.012) | (0.118) |
| Mean outside coverage | 0.415 | 0.326 | 0.170 | 0.012 | 0.143 |
| Bandwidth (dBm) | 13.02 | 10.94 | 13.50 | 10.47 | 14.34 |
| Observations | 445 | 331 | 445 | 331 | 477 |
| Villages | 146 | 111 | 146 | 111 | 166 |

*Notes:* Results present estimates of $\beta$ using a local linear regression specification of Equation (1) on information and misinformation. The observations are respondents in the survey sample (2016). Optimal bandwidth chosen as in Calonico et al. (2014). All specifications include controls for elevation, slope, sex, age, urban, and education level, and categories for Kpelle, Bassa, and other tribes. Standard errors clustered at the village level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

# Appendix A Estimation of the Irregular Terrain Model (ITM)

The estimation of cellphone coverage involved two major steps. First, we estimated the Irregular Terrain Model (ITM) to obtain signal strength on the ground for each individual tower using the CloudRF API (available at: `https://api.cloudrf.com`). Table A1 lays out the specific inputs used in the estimation of the ITM. For each tower, the output from the cloudRF API is an ArcGIS-readable polygon shapefile containing as attribute the signal received on the ground for different bands of the polygon file. For each tower, the estimated signal strength ranged between -50 and -140 dBm. We clarify that the cloudRF API actually gives the absolute value of the signal strength, thus technically the output ranged between 50 and 140.

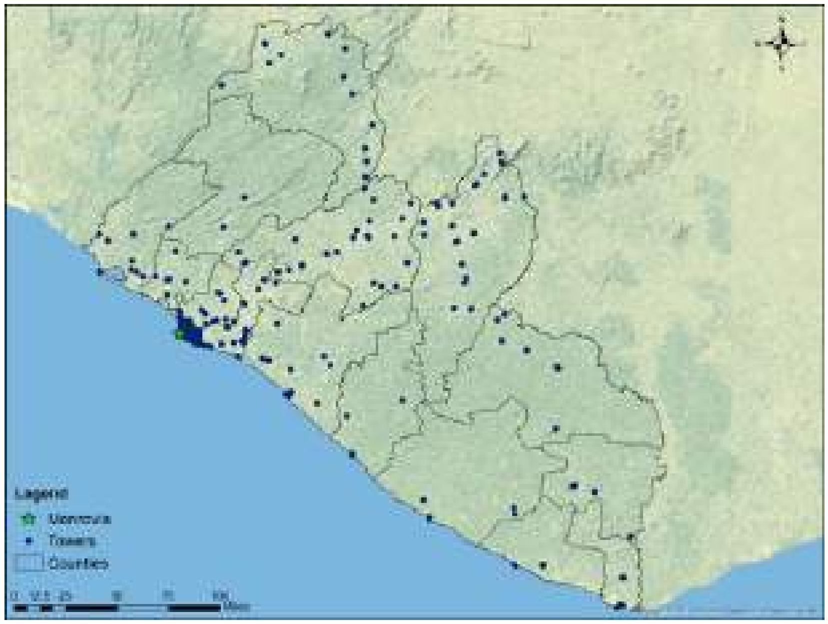
Second, the polygon shapefiles obtained from the ITM were processed in Arcmap 10.7 using the ArcPy package. The processing involved three steps. First, we projected the shapefiles using the *Project* tool to an appropriate coordinate system (UTM 29N) for analysis. Second, we converted the shapefiles to raster files with cell size of 50X50 meters using the *Polygon to Raster* tool. Third, we used the *Mosaic to Raster* tool to combine the raster files for each tower into one single raster file for all of Liberia (depicted in Figure 1). For raster cells where there was overlap in coverage between several towers, we used the "Minimum" rule to assign the signal strength to that particular cell. This assigns the highest signal strength available to that cell since the lower the estimated absolute value dBm, the higher the signal strength.

Table A1: Main Variables and Parameters of Irregular Terrain Model (ITM)

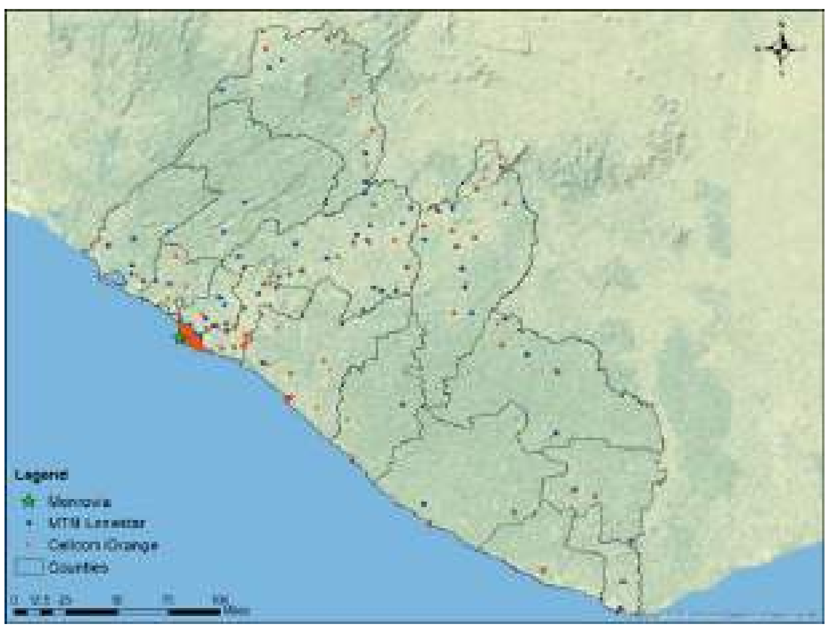
| Model variables | Description | Parameters |
|---|---|---|
| *Transmitter (cell tower) characteristics* | | |
| Transmitter power (txw) | Transmission power (Watts) | 5 |
| Frequency (frq) | Radio wave frequency (MHz) | 900 |
| Latitude (lat) | Latitude of cell tower | e.g., 6.27069 |
| Longitude (lon) | Longitude of cell tower | e.g., -10.73158 |
| Transmitter height (txh) | Height of cell tower above ground (meters) | 30 |
| Radius (rad) | Maximum coverage radius (kilometers) | 20 |
| Antenna gain (txg) | Transmitter antenna gain (dBi) | 2.14 |
| Antenna (ant) | Antenna type | 39 (omnidirectional) |
| Azimuth (azi) | Antenna azimuth angle (degrees) | 0° |
| Antenna tilt (tlt) | Antenna vertical angle (degrees) | 0° |
| Antenna polarization (pol) | Vertical/Horizontal | v (vertical) |
| *Receiver (handheld device) characteristics* | | |
| Receiver sensitivity (rxs) | Minimum power received threshold (dBm) | -140 |
| Receiver height (rxh) | Receiver height above ground (meters) | 1.5 |
| Antenna gain (rxg) | Receiver antenna gain (dBi) | 2.14 |
| *Geographic characteristics* | | |
| Resolution (res) | Topographic model | 30 (DSM30) |
| Clutter (clh) | Consider clutter (trees, buildings, etc.) | Yes |
| Climate (cli) | 1: Equatorial | 3 |
| | 2: Continental Subtropical | |
| | 3: Maritime Subtropical | |
| | 4: Desert | |
| | 5: Continental Temperate | |
| | 6: Maritime Temperate, over land | |
| | 7: Maritime Temperate, over sea | |
| Terrain conductivity (ter) | Salt water : 80 | 15 (farmland) |
| | Fresh water : 80 | |
| | Good ground : 25 | |
| | Marshy land : 12 | |
| | Farmland, forest : 15 | |
| | Average ground : 15 | |
| | Mountain, sand : 13 | |
| | City : 5 | |
| | Poor ground : 4 | |
| *Other inputs* | | |
| Model (pm) | Propagation model | 1 (ITM) |
| Model subtype (pe) | Conservative, Average, Optimistic | 2 (Average) |
| Measure (out) | Measurement units | 2 (dBm) |
| Engine (eng) | API Engine | 2 (SLEIPNIR) |
| Knife edge diffraction (ked) | Yes(=1) or No(=0) | 0 |
| Color scheme (col) | Color scheme | 9 (Grayscale/GIS) |

*Notes:* Refer to CloudRF API (available at: https://api.cloudrf.com for more information. CloudRF API code within parenthesis in "Model variables" column.

## Appendix B Additional Figures and Tables




(a) All Towers




(b) By Network Operator

Figure B1: Cellphone Towers, Liberia (2013)

*Notes:* Cell towers' location obtained from the Liberia Telecommunications Authority (LTA).

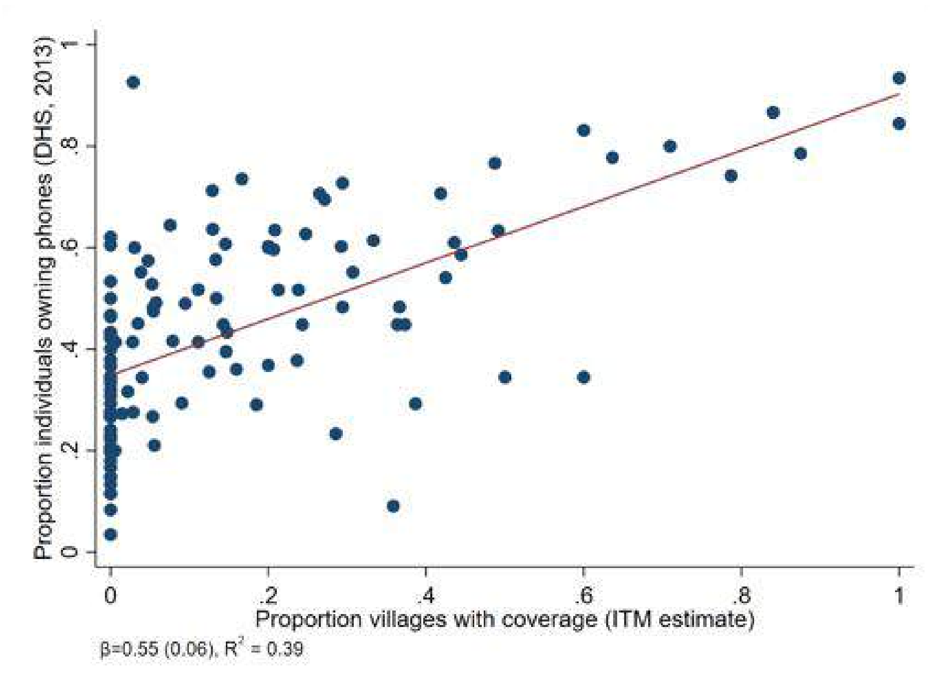
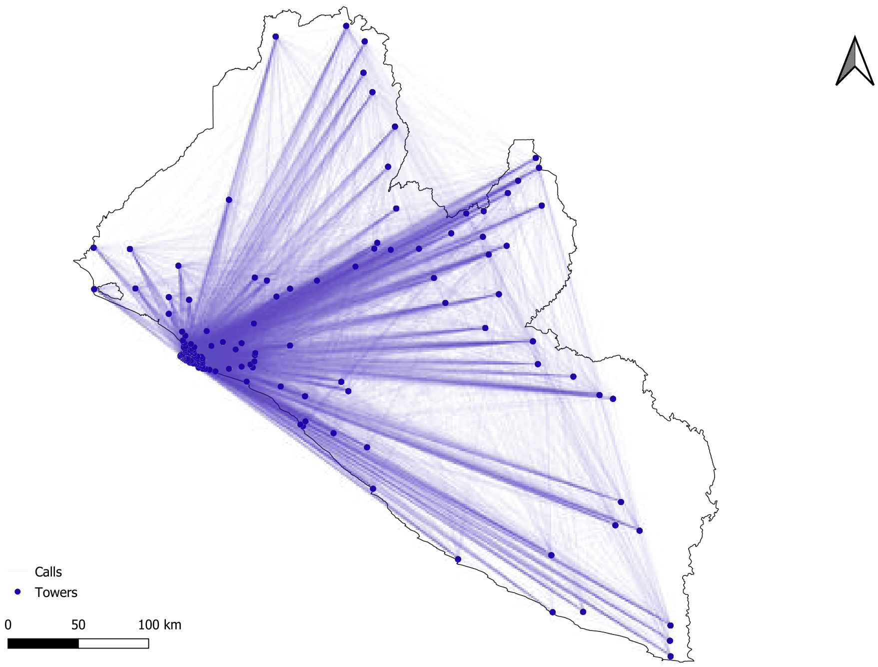
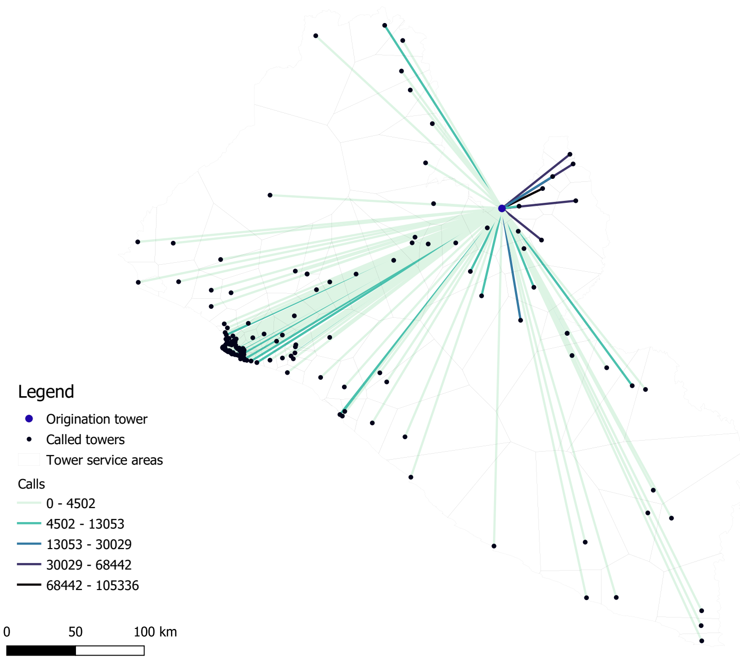



Figure B2: Predicted Coverage *versus* Reported Cellphone Ownership

*Notes*: Correlation between the proportion of villages within a district that are estimated to have cellphone coverage by the ITM (X axis) and the proportion of individuals reporting owning a cellphone in the Demographic and Health Survey (DHS, 2013) (Y axis).



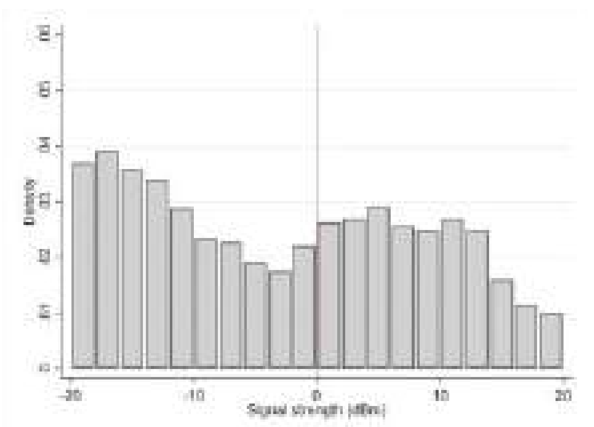

(a) All calls (June 15, 2015)



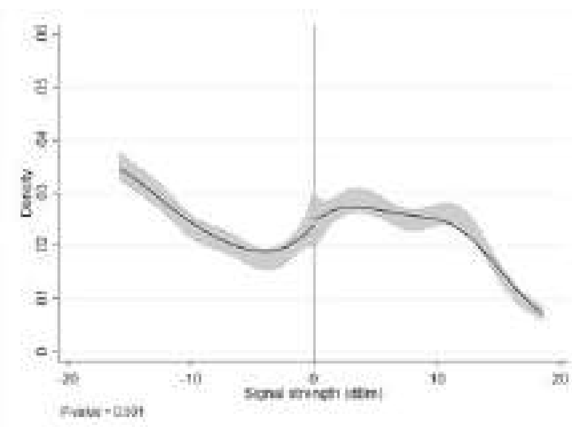

(b) Sample calls and called towers

Figure B3: Example of Call Detail Records

*Notes:* Panel a: All calls for the day of June 15, 2013. Panel b: Calls originating from cell tower in Ganta city (highlighted in blue) to all other towers in Liberia (black dots) in June and July of 2015. Color gradient represents the number of calls originating from Ganta city to the indicated tower. Darker shades mean a larger number of calls going to the indicated tower. Polygons indicate the service areas for each tower. Service areas calculated using a Voronoi/Thiessen polygon.



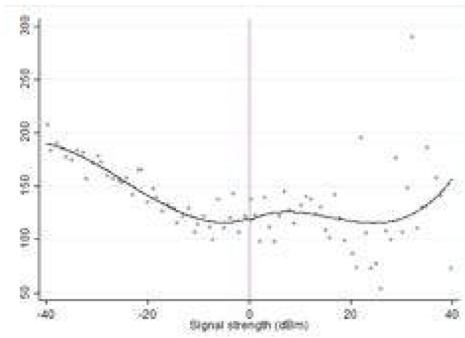

(a) Histogram



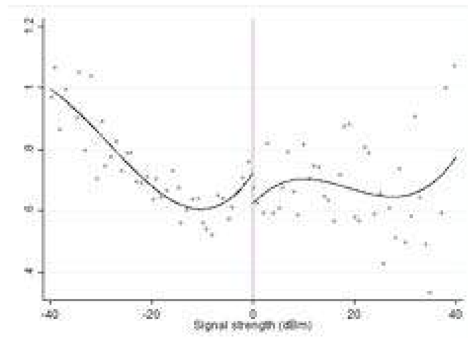

(b) Test for manipulation

Figure B4: Histogram and Test for Manipulation of Forcing Variable

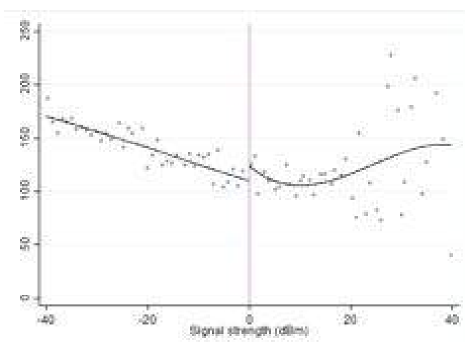
*Notes:* "Signal strength (dBm)" refers to the received power or signal strength, measured in decibel milliwatts, at the village level. Signal strength is normalized so that negative values mean no coverage (i.e., not enough signal strength). Histogram bar width is 2 dBm. Panel (b) uses the test for breaks in the density of the forcing variable proposed in Cattaneo et al. (2019) and uses the code discussed in Cattaneo et al. (2018). P-value for test presented in figure caption.



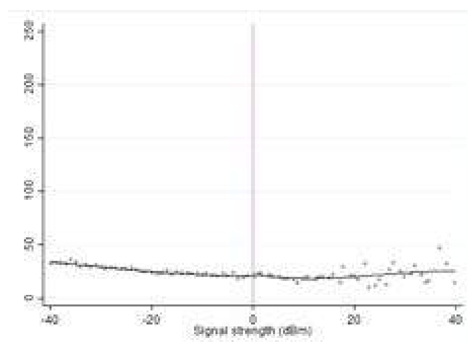

Elevation (m)




Slope (%)




Distance to Monrovia (km)




Distance to closest city (km)

Figure B5: Regression Discontinuity Plots, Covariates

*Notes:* Solid dots give the average of the specified variable for each bin of signal strength (dBm). “Signal strength (dBm)” refers to the received power or signal strength, measured in decibel milliwatts, at the village level. Signal strength is normalized so that negative values mean no coverage (i.e., not enough signal strength). The solid line trends give the predicted values from a regression of the outcome variable on a fourth degree polynomial in distance to the boundary that uses a triangular kernel.

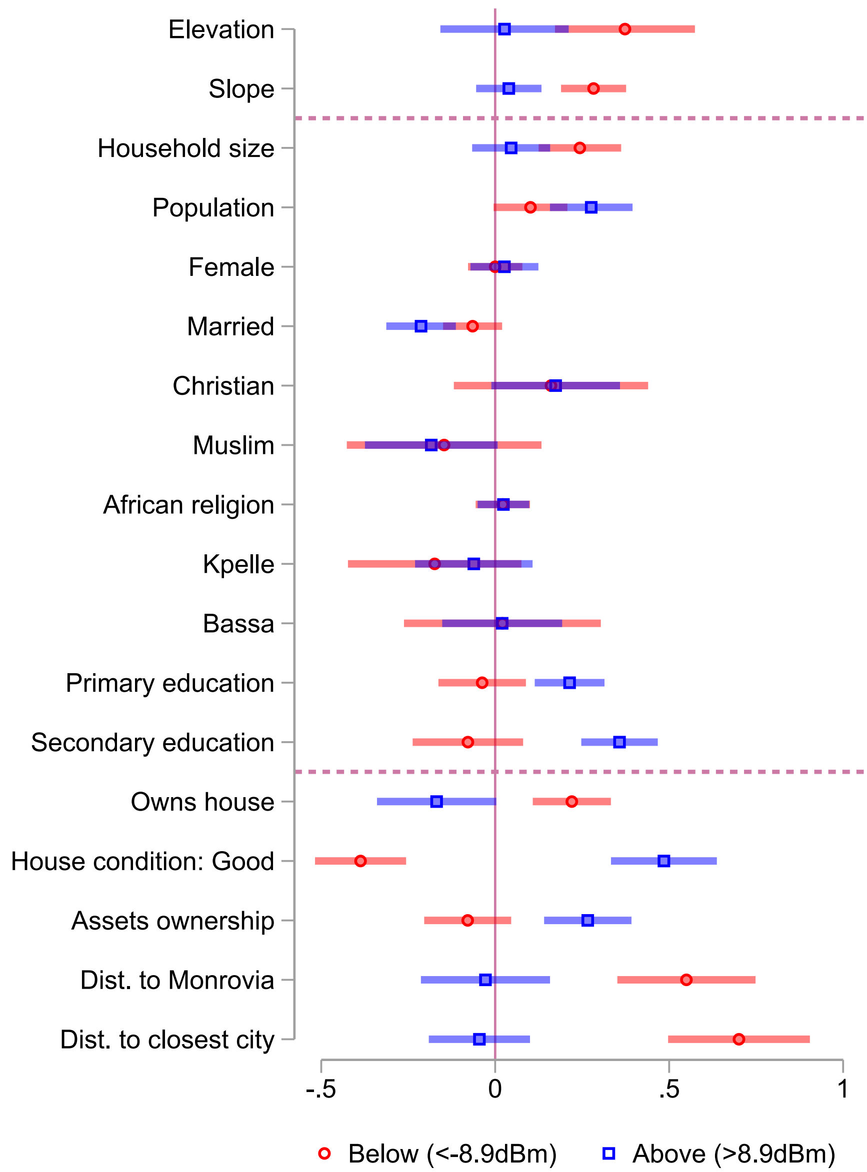



Figure B6: Difference in Means (RD Sample Villages vs Other Villages)

*Notes:* All variables are standardized for convenience. For each variable, dots give the difference in means between the specified group and the RD estimation sample. RD estimation sample is the sample within the Calonico et al. (2014) estimated bandwidth of analysis: [-8.9dBm, 8.9dBm]. Red dots refer to the sample below -8.9 dBm (less coverage than RD estimation sample). Blue dots refer to the sample above 8.9 dBm (more coverage than RD estimation sample). Lines around the dots give the 95% confidence interval for the difference in means. Dashed horizontal lines divide variables by topographic, demographic, and economic characteristics.

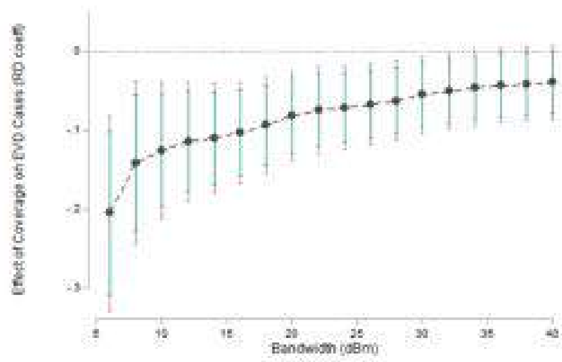



Figure B7: Bandwidth Sensitivity

*Notes:* Solid dots indicate the RD estimate from Equation (1) using the specified bandwidth. Range spikes indicate 95% and 90% confidence intervals of the estimates.

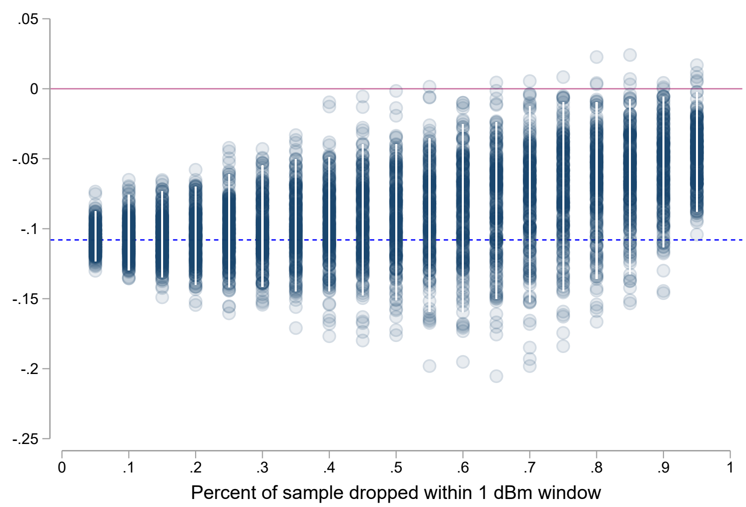



Figure B8: Sensitivity of Results to Observations near Treatment Cutoff

*Notes*: Each dot gives the estimated RD coefficient from 250 replications that estimate Equation (1) after randomly dropping the specified percent of observations within a 1 dBm window around the cutoff. White vertical lines indicate the 2.5th and 97.5th percentile of the estimated 250 coefficients within each percent dropped category. Blue dashed line gives the RD coefficient estimate from estimating Equation (1) without any restrictions (i.e., dropping 0% of observations near cutoff).

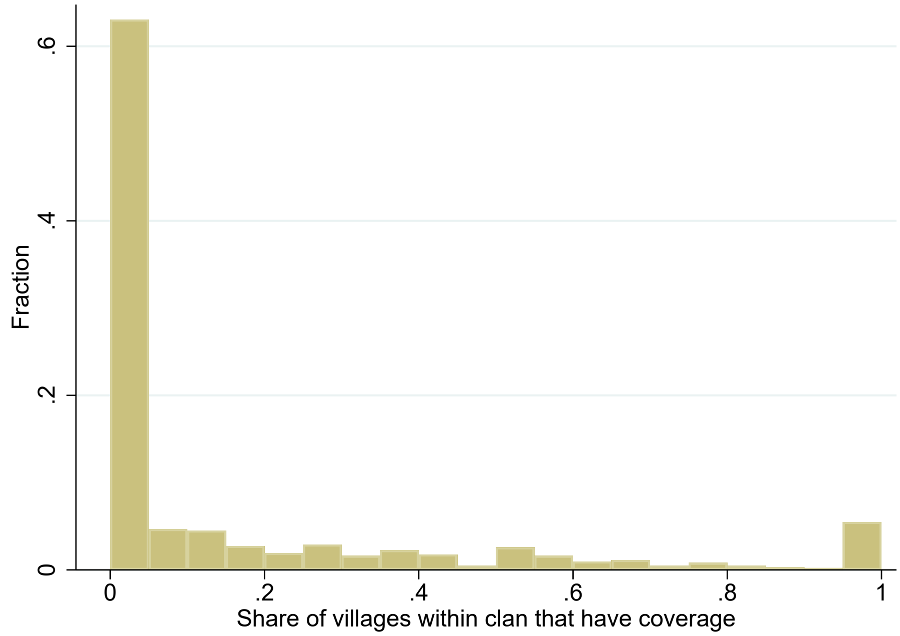
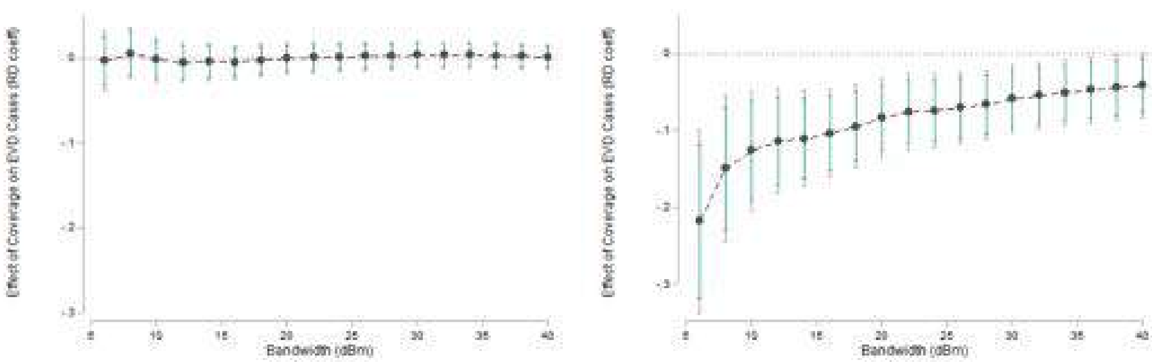



Figure B9: Histogram of the share of villages within a clan that have coverage

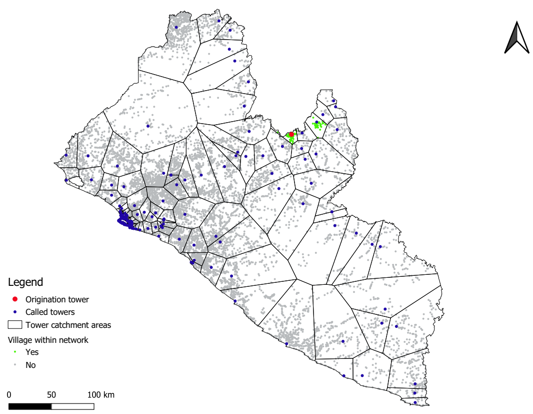



(a) Pre-hotline

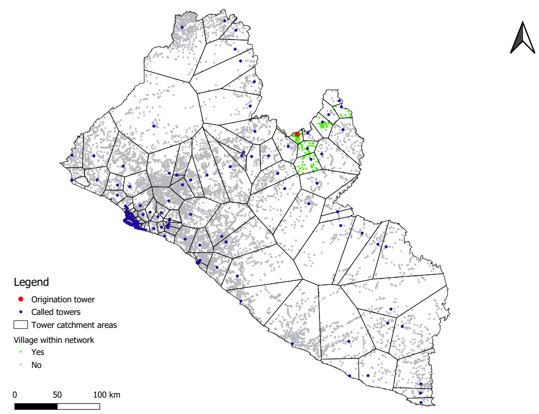
(b) Post-hotline

Figure B10: Bandwidth Sensitivity, by Hotline Timing

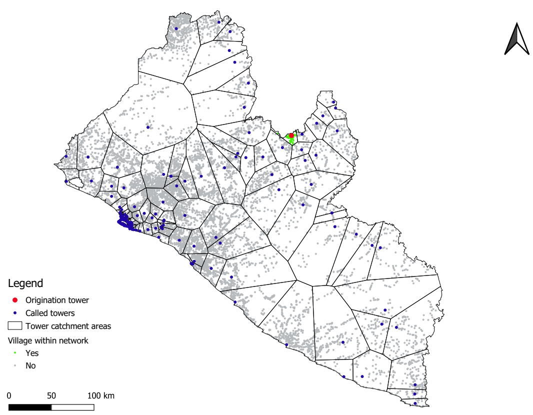
*Notes:* Solid dots indicate the RD estimates from Equation (1) using the specified bandwidth. Range spikes indicate 95% and 90% confidence intervals of the estimates.



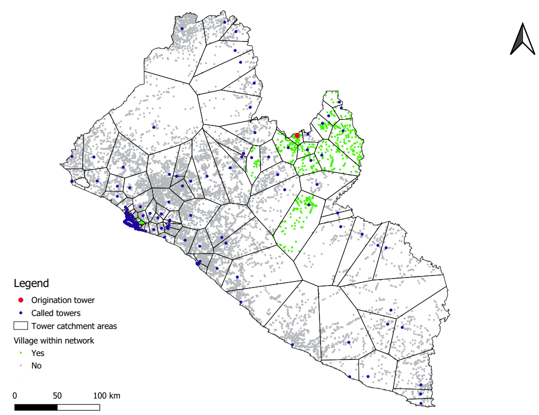

(a) Top tower



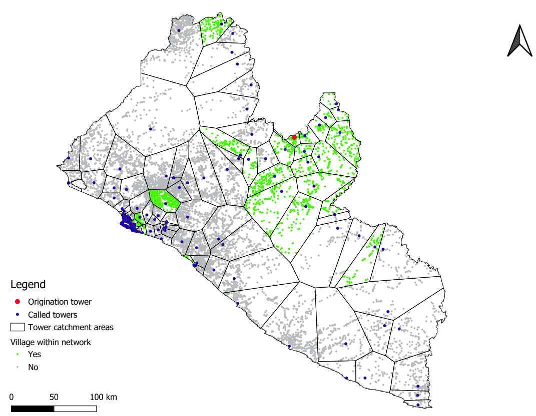

(b) Top 5 towers




(c) Top 50% calls



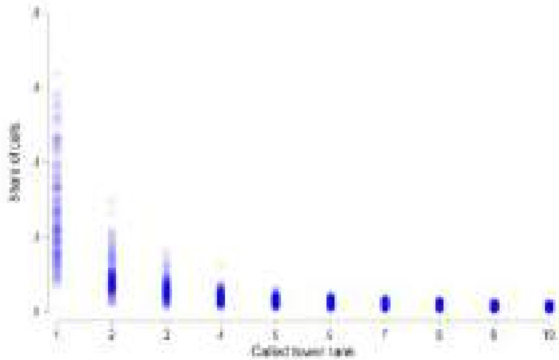

(d) Top 75% calls



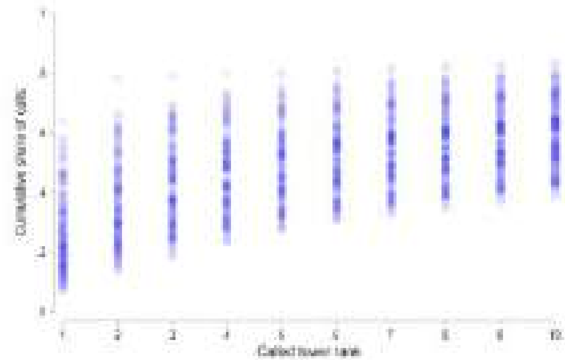

(e) Top 90% calls

Figure B11: Example of Network for Ganta City, Different Definitions

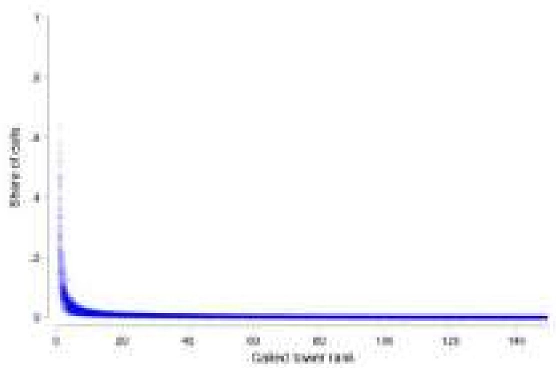
*Notes:* Networks for Ganta city tower, different definitions. Villages that are part of the network according to the specified definition are highlighted. Villages outside the network are gray. Given tower $k$, panels a and b define a network as all villages within the service area of tower $k$ and the top and top-5 towers receiving calls originating from $k$. Panels c-e define a network as all villages within the service area of tower $k$ and the towers receiving 50%, 75%, and 90%, respectively, of all calls originating from $k$. CDR data obtained from Cellcom Liberia.



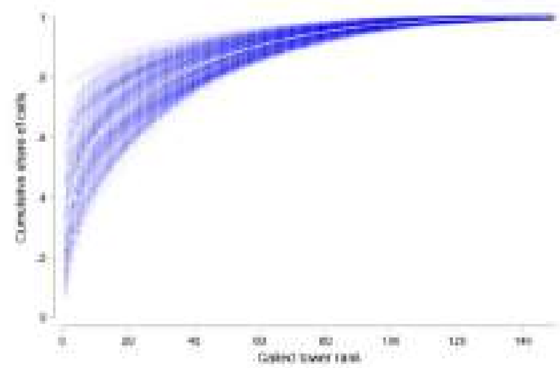
(a) Share of calls per tower (top-10 towers)



(b) Cumulative share of calls per tower (top-10 towers)



(c) Share of calls per tower



(d) Cumulative share of calls per tower

Figure B12: Share and Cumulative Share of Calls per Tower

*Notes:* Panels (a) and (c) depict for all towers in the sample, the share of all outgoing calls by the rank of the tower called. For example, the share of outgoing calls going to the top-ranked tower (most called tower) ranged between 10 to slightly above 60% (panel a). Panels (b) and (d) depict the cumulative share of outgoing calls.

Table B1: Summary Statistics by Coverage Status (Villages)

| | Full sample | | | Within 20 dBm | | | Within 10 dBm | | |
|---|---|---|---|---|---|---|---|---|---|
| | Inside (1) | Outside (2) | S.E. (3) | Inside (4) | Outside (5) | S.E. (6) | Inside (7) | Outside (8) | S.E. (9) |
| *Topographic characteristics:* | | | | | | | | | |
| Elevation (m) | 122.8 | 165.3 | (14.69)*** | 123.5 | 124.6 | (12.26) | 122.3 | 117.7 | (11.69) |
| Slope (%) | 0.68 | 0.89 | (0.05)*** | 0.68 | 0.64 | (0.04) | 0.67 | 0.62 | (0.04) |
| *Demographic characteristics:* | | | | | | | | | |
| Household size | 4.68 | 5.02 | (0.10)*** | 4.66 | 4.58 | (0.07) | 4.64 | 4.52 | (0.08) |
| Population (log) | 4.38 | 4.32 | (0.07) | 4.25 | 4.17 | (0.06) | 4.16 | 4.19 | (0.08) |
| Female (%) | 0.48 | 0.48 | (0.00) | 0.48 | 0.48 | (0.00) | 0.48 | 0.48 | (0.01) |
| Married (%) | 0.37 | 0.38 | (0.01) | 0.37 | 0.39 | (0.01)** | 0.38 | 0.39 | (0.01) |
| Christian (%) | 0.85 | 0.87 | (0.03) | 0.85 | 0.85 | (0.02) | 0.83 | 0.85 | (0.02) |
| Muslim (%) | 0.11 | 0.10 | (0.03) | 0.11 | 0.13 | (0.02) | 0.13 | 0.13 | (0.02) |
| African religion (%) | 0.01 | 0.01 | (0.00) | 0.01 | 0.00 | (0.00) | 0.00 | 0.00 | (0.00) |
| Kpelle (%) | 0.31 | 0.28 | (0.04) | 0.31 | 0.33 | (0.03) | 0.31 | 0.36 | (0.03)* |
| Bassa (%) | 0.25 | 0.24 | (0.06) | 0.26 | 0.24 | (0.04) | 0.26 | 0.21 | (0.04) |
| Other ethnic group (%) | 0.43 | 0.48 | (0.05) | 0.42 | 0.42 | (0.03) | 0.42 | 0.42 | (0.03) |
| *Economic characteristics:* | | | | | | | | | |
| Primary education (%) | 0.28 | 0.25 | (0.01)** | 0.28 | 0.26 | (0.01)** | 0.26 | 0.26 | (0.01) |
| Secondary education (%) | 0.38 | 0.33 | (0.02)*** | 0.37 | 0.34 | (0.01)** | 0.35 | 0.35 | (0.01) |
| Owns house (%) | 0.80 | 0.87 | (0.02)*** | 0.81 | 0.84 | (0.01)*** | 0.82 | 0.82 | (0.01) |
| House condition: Good (%) | 0.26 | 0.14 | (0.02)*** | 0.25 | 0.20 | (0.01)*** | 0.22 | 0.21 | (0.01) |
| Asset ownership (%) | 0.13 | 0.11 | (0.01)*** | 0.13 | 0.12 | (0.01)* | 0.12 | 0.12 | (0.01) |
| Distance to Monrovia (km) | 112.3 | 166.9 | (12.51)*** | 111.1 | 126.1 | (10.08) | 111.3 | 119.9 | (9.11) |
| Distance to closest city (km) | 19.82 | 33.72 | (2.15)*** | 19.67 | 22.52 | (1.36)** | 19.58 | 21.06 | (1.24) |
| Observations | 1,856 | 7,830 | | 1,698 | 2,141 | | 1,112 | 801 | |

*Notes:* Columns (1), (2), (4), (5), (7) and (8) give the means of the corresponding variable. Columns (3), (6) and (9) give the clustered standard errors for the difference in means in parenthesis. Clustered standard errors at the district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table B2: Summary Statistics, RD Sample Villages *versus* Other Villages

| | (1) Within Bandwith $[-8.9dBm, 8.9dBm]$ | (2) Below $(< -8.9dBm)$ | (3) Above $(> 8.9dBm)$ | (4) P-value ((1)-(2)) | (5) P-value ((1)-(3)) |
|---|---|---|---|---|---|
| *Topographic characteristics:* | | | | | |
| Elevation (m) | 120.10 | 169.42 | 123.65 | 0.000 | 0.775 |
| | (119.46) | (131.93) | (137.61) | | |
| Slope (%) | 0.66 | 0.91 | 0.69 | 0.000 | 0.413 |
| | (0.62) | (0.94) | (0.67) | | |
| *Demographic characteristics:* | | | | | |
| Household size | 4.62 | 5.06 | 4.70 | 0.000 | 0.422 |
| | (1.52) | (1.85) | (1.59) | | |
| Population (log) | 4.18 | 4.33 | 4.58 | 0.061 | 0.000 |
| | (1.40) | (1.41) | (1.77) | | |
| Female (%) | 0.48 | 0.48 | 0.49 | 1.000 | 0.593 |
| | (0.10) | (0.09) | (0.09) | | |
| Married (%) | 0.39 | 0.38 | 0.36 | 0.133 | 0.000 |
| | (0.12) | (0.11) | (0.11) | | |
| Christian (%) | 0.83 | 0.88 | 0.88 | 0.258 | 0.067 |
| | (0.30) | (0.25) | (0.24) | | |
| Muslim (%) | 0.13 | 0.10 | 0.09 | 0.302 | 0.061 |
| | (0.29) | (0.24) | (0.22) | | |
| African religion (%) | 0.00 | 0.01 | 0.01 | 0.579 | 0.530 |
| | (0.03) | (0.03) | (0.03) | | |
| Kpelle (%) | 0.34 | 0.27 | 0.31 | 0.170 | 0.475 |
| | (0.40) | (0.39) | (0.36) | | |
| Bassa (%) | 0.24 | 0.24 | 0.24 | 0.885 | 0.819 |
| | (0.38) | (0.40) | (0.37) | | |
| Other ethnic group (%) | 0.42 | 0.48 | 0.44 | 0.264 | 0.627 |
| | (0.42) | (0.45) | (0.40) | | |
| Primary education (%) | 0.26 | 0.25 | 0.30 | 0.556 | 0.000 |
| | (0.17) | (0.18) | (0.16) | | |
| Secondary education (%) | 0.35 | 0.33 | 0.42 | 0.330 | 0.000 |
| | (0.21) | (0.21) | (0.22) | | |
| *Economic characteristics:* | | | | | |
| Owns house (%) | 0.82 | 0.87 | 0.78 | 0.000 | 0.056 |
| | (0.27) | (0.22) | (0.29) | | |
| House condition: Good (%) | 0.21 | 0.14 | 0.31 | 0.000 | 0.000 |
| | (0.21) | (0.17) | (0.27) | | |
| Asset ownership (%) | 0.12 | 0.11 | 0.15 | 0.213 | 0.000 |
| | (0.09) | (0.09) | (0.11) | | |
| Distance to Monrovia (km) | 114.47 | 171.30 | 111.58 | 0.000 | 0.767 |
| | (89.33) | (102.05) | (103.47) | | |
| Distance to closest city (km) | 20.41 | 34.84 | 19.47 | 0.000 | 0.540 |
| | (15.61) | (20.67) | (16.42) | | |
| Observations | 1547 | 7207 | 932 | | |

*Notes:* Columns (1), (2), (3) give the means and standard deviations of the corresponding variable. Columns (4) and (5) give the p-value for the difference in means between column 1 and 2 (column 4) and columns 1 and 3 (column 5). Clustered standard errors at the district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table B3: Effect of Coverage on Likelihood of EVD Suspected Death

| | Dep. Variable = $\mathbb{1}\{\text{Number of Suspected EVD Deaths} > 0\}$ | | | | | |
|---|---|---|---|---|---|---|
| | Baseline | Controls | Topography Poly | Polynomial RD | Kernel choice | Probit model |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Coverage | -0.069* | -0.074** | -0.069* | -0.058 | -0.067* | -0.243* |
| | (0.037) | (0.030) | (0.036) | (0.036) | (0.035) | (0.129) |
| Mean outside coverage | 0.111 | 0.111 | 0.109 | 0.091 | 0.106 | 0.087 |
| Bandwidth (dBm) | 10.64 | 10.26 | 11.10 | 50.00 | 9.02 | . |
| Observations | 1913 | 1913 | 2139 | 7014 | 1741 | 1741 |
| Districts | 86 | 86 | 87 | 115 | 84 | 84 |
| Marginal Effect | | | | | | -0.048 |

*Notes:* Columns (1) and (2) present estimates of $\beta$ using a local linear regression specification of Equation (1). All specification include controls for elevation and slope. Column (2) adds controls for average household size, population, female, married, Christian, Muslim, African religion, Kpelle, Bassa, primary education, literacy, house ownership and quality, an index of asset ownership, distance to Monrovia, and distance to closest main city. Optimal bandwidth chosen as in Calonico et al. (2014). Column (3) uses topography polynomial: $h(\mathbf{G}_i) = \rho_1 elev_i + \rho_2 elev_i^2 + \rho_3 slope_i + \rho_4 elev_i \times slope_i + \rho_5 elev_i^2 \times slope_i + \rho_6 slope_i^2 + \rho_7 elev_i \times slope_i^2 + \rho_8 elev_i^2 \times slope_i^2$. It also includes distance to tower as additional control. Column (4) uses a wider bandwidth of 50 dBm and a third degree polynomial specification for $f(\tilde{R}_i)$. Column (5) uses a rectangular kernel. Column (6) estimates a probit specification of Equation (1). Optimal bandwidths chosen as in Calonico et al. (2014) except for column (4) which uses a fixed, wider bandwidth. Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table B4: Effect of Coverage on Number of Months Affected by EVD

| | Number of Months Affected by EVD | | | | |
|---|---|---|---|---|---|
| | Baseline | Controls | Topography Poly | Polynomial RD | Kernel choice |
| | (1) | (2) | (3) | (4) | (5) |
| Coverage | -0.209** | -0.188** | -0.220** | -0.158** | -0.199** |
| | (0.096) | (0.084) | (0.099) | (0.074) | (0.087) |
| Mean outside coverage | 0.128 | 0.128 | 0.128 | 0.090 | 0.142 |
| Bandwidth (dBm) | 8.72 | 8.97 | 8.59 | 50.00 | 7.10 |
| Observations | 1547 | 1547 | 1547 | 7014 | 1369 |
| Districts | 83 | 83 | 83 | 115 | 81 |

*Notes:* Columns (1) and (2) present estimates of $\beta$ using a local linear regression specification of Equation (1). All specification include controls for elevation and slope. Column (2) adds controls for average household size, population, female, married, Christian, Muslim, African religion, Kpelle, Bassa, primary education, literacy, house ownership and quality, an index of asset ownership, distance to Monrovia, and distance to closest main city. Optimal bandwidth chosen as in Calonico et al. (2014). Column (3) uses topography polynomial: $h(\mathbf{G}_i) = \rho_1 elev_i + \rho_2 elev_i^2 + \rho_3 slope_i + \rho_4 elev_i \times slope_i + \rho_5 elev_i^2 \times slope_i + \rho_6 slope_i^2 + \rho_7 elev_i \times slope_i^2 + \rho_8 elev_i^2 \times slope_i^2$. t also includes distance to tower as additional control. Column (4) uses a wider bandwidth of 50 dBm and a third degree polynomial specification for $f(\tilde{R}_i)$. Column (5) uses a rectangular kernel. Optimal bandwidths chosen as in Calonico et al. (2014) except for column (4) which uses a fixed, wider bandwidth. Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table B5: Effect of Coverage on Likelihood of EVD Case, By Distance to Coverage Boundary

| | Geographic Distance to Coverage Boundary | | | | | |
|---|---|---|---|---|---|---|
| | No restriction | Within 2km | Within 4km | No restriction | Within 2km | Within 4km |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Coverage | -0.108** | -0.110** | -0.107** | -0.101** | -0.105** | -0.100** |
| | (0.048) | (0.049) | (0.049) | (0.042) | (0.043) | (0.042) |
| Mean outside coverage | 0.093 | 0.094 | 0.093 | 0.093 | 0.093 | 0.093 |
| Bandwidth (dBm) | 8.99 | 9.04 | 8.96 | 8.13 | 8.15 | 8.07 |
| Observations | 1547 | 1729 | 1542 | 1547 | 1535 | 1542 |
| Districts | 83 | 84 | 83 | 83 | 83 | 83 |
| Controls | No | No | No | Yes | Yes | Yes |

*Notes:* Dependent variable equal 1 if there is at least one EVD case reported within the village. Columns (1)-(3) present estimates of $\beta$ using a local linear regression specification of Equation (1). All specification include controls for elevation and slope. Columns (4)-(6) add controls for average household size, population, female, married, Christian, Muslim, African religion, Kpelle, Bassa, primary education, literacy, house ownership and quality, an index of asset ownership, distance to Monrovia, and distance to closest main city. Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table B6: Average Travel Distance to Closest Village with Coverage

| | No dBm restrictions (1) | Within 20 dBm (2) | Within 10 dBm (3) |
|---|---|---|---|
| Mean (km) | 25.94 | 6.73 | 5.69 |
| Std. Dev. | 45.69 | 10.45 | 10.26 |
| Observations | 4,639 | 1,060 | 367 |
| Share of non-coverage sample | 0.59 | 0.50 | 0.46 |

*Notes:* Stata program "georoute" (Weber and Péclat, 2017) used for travel distance calculations. Program uses Open Source Routing Machine (OSRM) engine for calculations and OpenStreetMap as base map. Villages with calculated travel distance shorter than Euclidean distance dropped from analysis. Calculated travel distance can be shorter than Euclidean distance if village is far from road network used in the calculation and program uses segment of road available. Share of non-coverage sample refers to the share of non-coverage villages for which a travel distance was calculated.

Table B7: Effect of Coverage on Likelihood of EVD Case, Control Villages close to Coverage

| | Closest coverage village within: | | | | |
|---|---|---|---|---|---|
| | 2km (1) | 4km (2) | 6km (3) | 8km (4) | 10km (5) |
| Coverage | -0.114* | -0.098* | -0.106** | -0.103** | -0.104** |
| | (0.060) | (0.055) | (0.052) | (0.049) | (0.048) |
| Obs | 1515 | 1612 | 1651 | 1677 | 1695 |
| Bandwidth | 9.082 | 9.216 | 9.297 | 9.348 | 9.507 |

*Notes:* Estimates of $\beta$ using a local linear regression specification of Equation (1). All specifications include controls for elevation and slope. Optimal bandwidth chosen as in Calonico et al. (2014). Standard errors clustered at district level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively. Stata program "georoute" (Weber and Péclat, 2017) used for travel distance calculations. Program uses Open Source Routing Machine (OSRM) engine for calculations and OpenStreetMap as base map. Villages with calculated travel distance shorter than Euclidean distance dropped from analysis. Calculated travel distance can be shorter than Euclidean distance if village is far from road network used in the calculation and program uses segment of road available.

Table B8: Summary Statistics by Coverage Status (Survey Sample)

| | Full sample | | | Within 20 dBm | | | Within 10 dBm | | |
|---|---|---|---|---|---|---|---|---|---|
| | Inside | Outside | S.E. | Inside | Outside | S.E. | Inside | Outside | S.E. |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| *Topographic characteristics:* | | | | | | | | | |
| Elevation (m) | 191.1 | 227.2 | (42.99) | 189.8 | 157.9 | (50.68) | 195.8 | 124.7 | (77.05) |
| Slope (%) | 0.70 | 0.88 | (0.17) | 0.61 | 0.74 | (0.11) | 0.65 | 0.62 | (0.13) |
| *Demographic characteristics:* | | | | | | | | | |
| Household size | 4.79 | 4.98 | (0.19) | 4.64 | 4.75 | (0.32) | 4.40 | 4.99 | (0.52) |
| Christian | 0.86 | 0.87 | (0.02) | 0.86 | 0.83 | (0.04) | 0.87 | 0.83 | (0.06) |
| Muslim | 0.10 | 0.09 | (0.02) | 0.10 | 0.12 | (0.04) | 0.07 | 0.13 | (0.05) |
| Female | 0.37 | 0.32 | (0.03)* | 0.32 | 0.34 | (0.04) | 0.28 | 0.28 | (0.06) |
| Age | 32.02 | 33.91 | (0.54)*** | 33.45 | 34.69 | (0.94) | 33.82 | 34.33 | (1.37) |
| Married | 0.44 | 0.57 | (0.03)*** | 0.53 | 0.58 | (0.05) | 0.57 | 0.53 | (0.08) |
| Kpelle | 0.25 | 0.33 | (0.08) | 0.17 | 0.35 | (0.06)*** | 0.15 | 0.34 | (0.09)** |
| Bassa | 0.09 | 0.13 | (0.03) | 0.17 | 0.16 | (0.08) | 0.23 | 0.15 | (0.13) |
| Mano | 0.16 | 0.09 | (0.06) | 0.19 | 0.09 | (0.07) | 0.24 | 0.10 | (0.10) |
| Other tribe | 0.50 | 0.44 | (0.06) | 0.47 | 0.40 | (0.07) | 0.38 | 0.41 | (0.08) |
| Urban | 0.80 | 0.50 | (0.05)*** | 0.63 | 0.46 | (0.08)** | 0.62 | 0.50 | (0.13) |
| *Economic characteristics:* | | | | | | | | | |
| Secondary education | 0.62 | 0.58 | (0.03) | 0.62 | 0.57 | (0.04) | 0.67 | 0.57 | (0.06)* |
| Works for wage | 0.17 | 0.19 | (0.02) | 0.17 | 0.20 | (0.03) | 0.20 | 0.20 | (0.06) |
| Lost job (2013) | 0.15 | 0.16 | (0.02) | 0.17 | 0.22 | (0.03) | 0.17 | 0.24 | (0.06) |
| Not working | 0.28 | 0.18 | (0.02)*** | 0.26 | 0.17 | (0.03)*** | 0.25 | 0.19 | (0.05) |
| Self-employed | 0.40 | 0.48 | (0.03)*** | 0.42 | 0.47 | (0.05) | 0.42 | 0.48 | (0.09) |
| Wealth index (<$20^{th}$ percentile) | 0.18 | 0.34 | (0.03)*** | 0.20 | 0.35 | (0.05)*** | 0.24 | 0.27 | (0.07) |
| Wealth index (20-$40^{th}$ percentile) | 0.28 | 0.23 | (0.02)* | 0.26 | 0.21 | (0.03) | 0.25 | 0.29 | (0.06) |
| Wealth index (40-$60^{th}$ percentile) | 0.12 | 0.10 | (0.01)* | 0.14 | 0.09 | (0.02)* | 0.14 | 0.09 | (0.04) |
| Wealth index (60-$80^{th}$ percentile) | 0.20 | 0.19 | (0.02) | 0.17 | 0.20 | (0.03) | 0.17 | 0.14 | (0.04) |
| Wealth index (>$80^{th}$ percentile) | 0.22 | 0.14 | (0.03)*** | 0.24 | 0.15 | (0.04)** | 0.19 | 0.21 | (0.06) |
| Distance to Monrovia (km) | 145.03 | 172.26 | (31.81) | 157.44 | 130.70 | (26.38) | 175.02 | 116.62 | (40.31) |
| Distance to closest city (km) | 6.57 | 29.82 | (2.69)*** | 15.86 | 23.09 | (5.08) | 11.61 | 22.34 | (4.53)** |
| Observations | 1,495 | 648 | | 516 | 202 | | 245 | 86 | |

*Notes:* Columns (1), (2), (4), (5), (7) and (8) give the means of the corresponding variable. Columns (3), (6) and (9) give the clustered standard errors for the difference in means in parenthesis. Clustered standard errors at the village level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table B9: Effect of Coverage on Preventive Care

| | Contact tracing | Explain EVD | Hygiene meetings | Bring info (Task-force) | Explain burials | Prevention material | Preventive care index |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Coverage | 0.127 | 0.187* | -0.172 | 0.212** | 0.122 | 0.054 | 0.062 |
| | (0.082) | (0.112) | (0.110) | (0.090) | (0.097) | (0.145) | (0.128) |
| Mean outside coverage | 0.194 | 0.874 | 0.472 | 0.282 | 0.174 | 0.745 | 0.040 |
| Bandwidth (dBm) | 7.99 | 12.25 | 19.18 | 8.68 | 11.69 | 13.58 | 16.18 |
| Observations | 251 | 346 | 589 | 296 | 328 | 404 | 518 |
| Villages | 77 | 139 | 228 | 86 | 122 | 146 | 189 |

*Notes:* Results present estimates of $\beta$ using a local linear regression specification of Equation (1) on preventive care variables. The observations are respondents in the survey sample (2016). Optimal bandwidth chosen as in Calonico et al. (2014). The information index is constructed following Kling et al. (2007). All specifications include controls for elevation and slope. Standard errors clustered at the village level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

Table B10: Effect of Coverage on Treatment Care

| | Take sick | Take dead | Ambulance on-time | CCCs within 10km | Treatment care index |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Coverage | 0.123*** | -0.022 | 0.217** | 0.494** | 0.345** |
| | (0.038) | (0.051) | (0.087) | (0.223) | (0.156) |
| Mean outside coverage | 0.100 | 0.070 | 0.802 | 0.477 | -0.036 |
| Bandwidth (dBm) | 10.76 | 11.59 | 10.75 | 10.43 | 8.22 |
| Observations | 300 | 328 | 331 | 331 | 268 |
| Villages | 111 | 122 | 111 | 111 | 86 |

*Notes:* Results present estimates of $\beta$ using a local linear regression specification of Equation (1) on treatment care variables. The observations are respondents in the survey sample (2016). Optimal bandwidth chosen as in Calonico et al. (2014). The care index is constructed following Kling et al. (2007). All specifications include controls for elevation and slope. Standard errors clustered at the village level. *, **, and *** indicate 10, 5, and 1 percent significance, respectively.

## Appendix C Determining the Coverage Cutoff

In principle, although one may know the typical range for the minimum required signal strength or sensitivity cutoff, the exact cutoff $c$ is generally unknown (Farahani, 2008). We employ three methods to determine potential cutoff(s).

First, we use the Difference in Kernels estimator described in Qiu (2011) and Porter and Yu (2015). We estimate equation (5) below for each potential signal strength cutoff $r$. In order to limit our search, we restrict values of $r$ to be within 80 to 110 dBm.

$$M(r) = \frac{1}{nh}\sum_{i=1}^{n} Y_i K_1\left(\frac{R_i - r}{h}\right) - \frac{1}{nh}\sum_{i=1}^{n} Y_i K_0\left(\frac{R_i - r}{h}\right) \tag{5}$$

where $R_i$ denotes the ITM-estimated received power (i.e., signal strength) in each village $i$. $Y_i$ is our outcome of interest. $K_1(.)$ and $K_0(.)$ are one-sided kernel estimators on the right and left side of potential cutoff $r$, respectively. We use a triangular kernel and a bandwidth $h$ of 2 dBms. Intuitively, this procedure compares the kernel-weighted average of our outcome on the right and left sides of point $r$. $M(r)$ should be large if $r$ is a cutoff point and small if $r$ is not a cutoff point (or, if there is no treatment effect). Figure C1 presents the estimates of $M(r)$ for different values of $r$. There is a clear and distinct jump in the difference in the kernels estimator at $r = 95$.

Second, we employ one of the strategies described in Card et al. (2008). Specifically, we estimate a simple relationship between our outcome $Y_i$ and the estimated signal strength $R_i$ for several potential cutoffs $r$ but within a fixed interval $[\underline{R}, \overline{R}]$:

$$Y_i = \alpha + \beta \mathbf{1}\left[R_i > r\right] + f(R_i) + \varepsilon_i \quad \text{with} \quad R_i \in [\underline{R}, \overline{R}] \tag{6}$$

where $f(.)$ is a function of the estimated signal strength $R_i$. We then choose potential cutoff $r$ based on the $r$ that yields the highest $R^2$. Intuitively, if there is a discontinuity at $r$, then any specification of (6) that uses a cutoff different than $r$ is misspecified. As before, we restrict the search to values of $r$ within 80 to 110 dBm and fix the interval $[\underline{R}, \overline{R}]$ to be within these values as well.

Figure C2 plots the $R^2$ for the specified cutoffs $r$ and for a linear, quadratic, cubic, and quartic specifications of $f(R_i)$. Note that maximum fit is reached when equation (6) uses

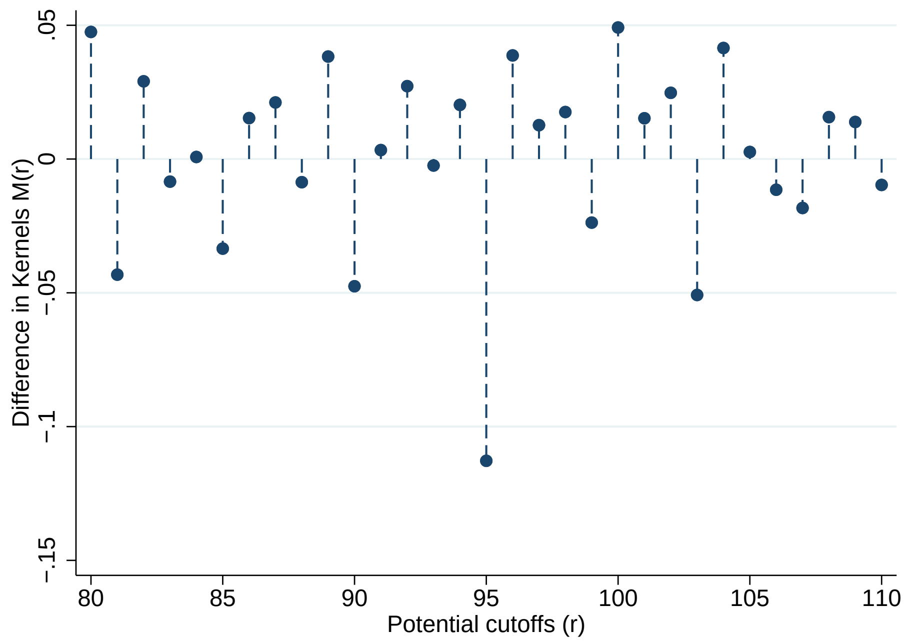



Figure C1: Difference in Kernels

*Notes:* Each dot represents the estimate of $M(r)$ in equation (5). Estimation

$r = 95$ suggesting that this is likely the cutoff at which our outcome jumps.

Lastly, we use a modification of the method proposed by Spokoiny (1998) adapted to our RD setting. Specifically, we proceed in three steps:

*Step 1:* We use a flexible polynomial $f(.)$ to estimate $Y_i = f(R_i - r_0) + \varepsilon_i$ within a neighborhood $U$ of point $r_0$. Given our setting, we start with $r_0 = 75$ which is to the left of any plausible cutoff point and a fourth degree polynomial specification for $f(.)$.

*Step 2:* We then examine the RMSE of the residuals from Step 1 at point $v \in V$ where $V = U \cap U'$ is the set of boundary points of interval $U$. In our one-dimensional case, $V$ are simply the left and right endpoints of the interval $U$. Furthermore, given that our potential cutoff is to the right of $r_0 = 75$, we focus only on the *right* endpoint residuals.

*Step 3:* We then gradually increase interval $U$ around $r_0$ and repeat steps 1-2 until there is a clear jump in the RMSE of endpoint residuals. Cutoff point $r$ is chosen as the endpoint of the maximal interval for which the endpoint residuals are "well-behaved" (i.e., the RMSE of the endpoint residuals does not jump significantly).

Intuitively, after we hit the maximal interval, the polynomial will have a hard time fitting

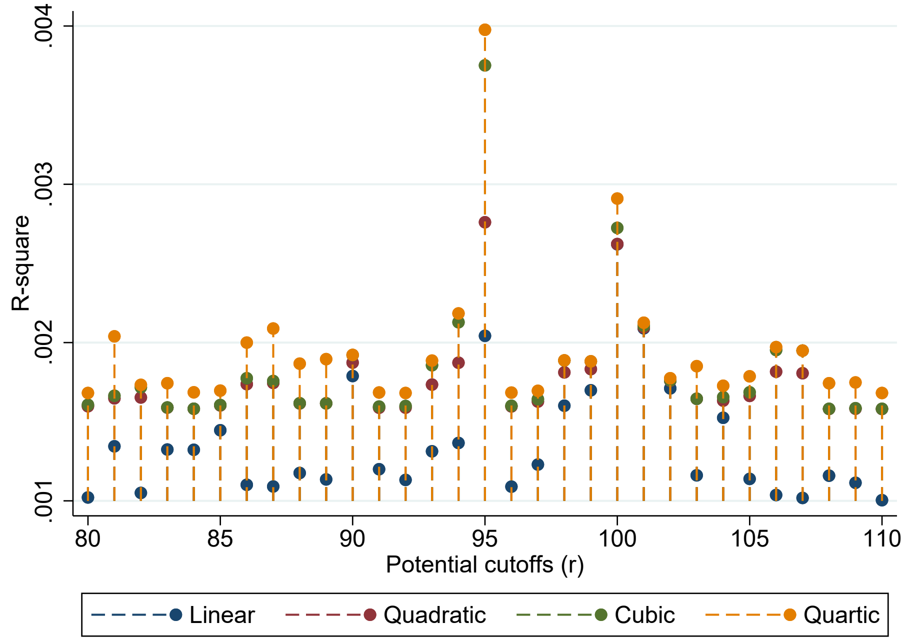
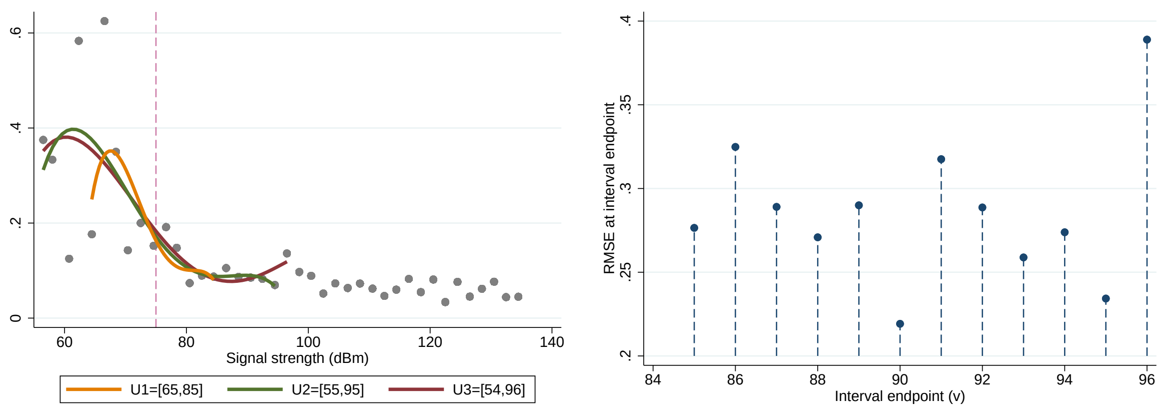



Figure C2: Model fit by Cutoff point $r$

*Notes:* Each dot represents the $R^2$ obtained from estimating equation (6) using different values for $r$ (x-axis), and using the specified polynomial for $f(R_i)$ in equation (6).

the jump in the outcome occurring at the cutoff point. Panel A of Figure C3 illustrates the method by presenting polynomial $f(R_i - r_0)$ estimated within three different intervals around $r_0 = 75$: intervals $U_1$, $U_2$, and $U_3$, with right endpoints at 85, 95, and 96 dBm, respectively. Note that the fit of $f(.)$ is reasonably well at the endpoints of $U_1$ and $U_2$, but performs poorly at the endpoint of $U_3$. This likely suggests that we have reached the maximal interval at $U_2$. This is confirmed in panel B of Figure C3 that plots the RMSE at the enpoints of each interval used to estimate $f(R_i - r_0)$. Notice that once we get to a right endpoint of 96, the RMSE increases substantially. This suggests that the cutoff point is 95, the right endpoint of the maximal interval.




(a) Polynomial within interval $U$ around $r_0 = 75$

(b) RMSE at interval endpoints

Figure C3: RMSE of Endpoint Residuals

*Notes:* Solid dots in Panel A give the average of the outcome variable for each bin of signal strength (dBm). "Signal strength (dBm)" refers to the received power or signal strength, measured in decibel milliwatts, at the village level. Bin width is 2 dBm. The solid lines are the polynomials estimated within the specified intervals. Dashed vertical line is the initial $r_0 = 75$. Refer to text for more detail.