

Pre-Analysis Plan

Turning Up, Tuning In, Turning Out: Experimental Evidence from Liberia¹

Jeremy Bowles² & Horacio Larreguy³

October 24, 2017

Abstract

Understanding the role of information is at the core of democratic accountability and the often-broken representative-constituent link. Representatives face strong incentives to reduce the dissemination of information about their policy promises to avoid tying their hands, and to avoid being held accountable for failing to meet them, especially so when the private returns to public office are high. When the media sector is underdeveloped then this results in a low-quality, low-information equilibrium where democratic accountability suffers. We evaluate the impact of an initiative designed to simultaneously shock the supply of programmatic information by candidates and the credibility of the media sector. We do this by leveraging experimental evidence from a nationwide debate initiative ahead of Liberia's 2017 elections for House of Representatives designed to solicit concrete policy promises from candidates. With random variation in the participation of political candidates and the intensity of debate broadcasting through community radio stations, we aim to parse how variation in exposure to the policy platforms of candidates affects levels of political information, voting behavior, electoral returns, and the role of the media in intermediating these effects. Ultimately, we want to assess whether the intervention was successful at breaking this low-quality, low-information equilibrium.

¹We are grateful to USAID for funding. With thanks to Internews Liberia, USAID/Liberia, the National Elections Commission of Liberia, NORC at the University of Chicago and GeoPoll for their cooperation. With special thanks to Jan McArthur for her enduring support, and Kate Thomas for research assistance. IRB approval granted by Harvard Committee on the Use of Human Subjects (IRB17-1178) and NORC at the University of Chicago (7554.072.01).

²PhD Candidate, Department of Government, Harvard University. jbowles@g.harvard.edu.

³Associate Professor, Department of Government, Harvard University. hlarreguy@fas.harvard.edu.

1 Overview

Democratic accountability relies on the effective selection of political candidates and the availability of mechanisms to monitor and incentivize them once in office. In contexts where the private returns to office-holding are high, corruption is entrenched and the link between representative and citizen fractured, understanding the mechanisms to improve accountability is a first-order concern. Following classic models of electoral behavior, much extant research focuses on the informational dimension of this problem: providing citizens with information affects both their selection of high-quality politicians and their ability to mitigate shirking once in office. However, focusing on the consequences of informational dissemination ignores the strategic calculus underlying its release. Politicians and political candidates in low-information environments face strong incentives to reduce the amount of information they provide about themselves – to avoid credibly committing themselves to implement particular policies and to avoid being monitored once in office. In developed countries, a functional media sector reduces the ability of politicians to control information in this way: either by supplying information about politicians directly or by reducing the flow of cheap talk from the politicians themselves by forcing them to commit to more credible promises. But when the media lacks credibility and capacity then low-quality, low-information equilibria persist, and democratic accountability suffers.

We suggest that – in weakly-institutionalized settings – one way to facilitate democratic accountability is to simultaneously target the supply-side decision of politicians in providing programmatic information about themselves, and the credibility of the media sector who are responsible for conveying this information. In partnership with USAID and Internews, we evaluate the impact of an initiative designed to study the consequences of a reform that took such an approach. Specifically, we randomize several elements of an initiative to hold debates between all 984 candidates for 73 House of Representatives seats ahead of the October 2017 election⁴. 129 standardized debates across all districts were designed to solicit the policy promises of different candidates in a setting where votes are most often won through purchase.

First, we randomize the encouragement to participate in the debates by varying the intensity of efforts to attract candidates to attend the debates. The decision to participate in a candidate debate is clearly a strategic one, and particularly so in clientelistic settings. Candidates who ‘win’ a debate may enjoy greater publicity and net electoral gains, but they risk either losing a debate or restricting their ability to deviate from policy promises on the campaign trail or once in office. Providing policy platforms through broadly-disseminated debates represents a shift from locally-disseminated cheap talk by candidates – promising to build schools, hospitals and roads everywhere – to a more costly signal of policy promises. Especially for leading candidates, the expected returns from debate participation are limited – they risk providing a platform for their challengers to attack them and gain publicity. Prominent examples abound of incumbents avoiding electoral debates: two examples from 2017 include Theresa May in the UK and

⁴We believe this is the first time debates have been held universally in an election in, at least, West Africa (Olukotun and Omotoso, 2017).

Uhuru Kenyatta in Kenya. Consistent with this calculus, other evaluations of debate initiatives have found that smaller competitors tend to benefit the most (Izama and Raffler, 2016).

Second, we randomize the intensity of radio coverage of the debates. Each debate is broadcast live by a community radio station, and in treatment districts debates are intensively rebroadcast ten times, at peak hours, at the height of the campaigning season. Across 43 community radio stations⁵, easily the dominant way to acquire political information in Liberia, we thus generate variation in the share of individuals exposed to candidate promises. Aside from affecting citizen information about participating candidates, this randomization also specifically affects the relative share of radio news focusing on programmatic policy in a context where candidates frequently turn radio stations into their own mouthpieces. As such, we consider that it may also affect perceptions of media bias and credibility.

With these two interventions, as well as several other sources of randomized variation in the administration of the debates, we evaluate a series of hypotheses. First, these focus on whether and how the initiative affected levels of political knowledge about the policy promises and competence of different candidates, as well as general information about policy. Second, on how learning about candidates affected candidate selection and the extent to which citizens vote in line with their preferences. Third, on the electoral returns to candidates and consequences for how candidates campaign. And fourth, on how debate exposure affects attitudes towards the media and the electoral process more broadly.

While fitting into the expansive literature on information and accountability, we contribute in several ways. First, while previous interventions has addressed the effect of localized debates (Bidwell, Casey and Glennerster, 2016; Izama and Raffler, 2016), we aim at assessing the effect of debate broadcasting. Differences in the modes of transmission might be great importance, as suggested by the mixed evidence from the plethora of studies studying the provision of information through leaflets and scorecards (Adena et al., 2015; Arias, 2016; Arias et al., 2017; Yanagizawa-Drott, 2014), with broadly more positive evidence on the consequences of media coverage of political issues (Ferraz and Finan, 2008; Larreguy, Marshall and Snyder, 2017). Modes of transmission that reach large shares of constituents – through which voters become aware that many other voters have also received a given piece of information – could produce powerful effects by inducing explicit or tacit voter coordination based on their common knowledge (Ferraz and Finan, 2008; Larreguy, Marshall and Snyder, 2017). Such coordination may in part explain the contrasting findings among the studies that look at the electoral effect of incumbent malfeasance revelations. Further, the broad media coverage of political information ought to undermine the ability of candidates to strategically buy votes in treated areas, and the intensity of radio broadcasting should reduce levels of political information decay found in similar interventions (Bidwell, Casey and Glennerster, 2016).

⁵The debates are broadcast by fewer than 73 radio stations since some have the ability to broadcast debates in more than one district. We explain the procedure to select radio stations for rebroadcasting in detail later.

2 Context

Liberia is among the world's poorest democracies. Since its emergence from civil war in 2003, the country has held two presidential (2005/2011), two House of Representatives (2005/2011) and two Senatorial elections (2005/2014). While public opinion data reveals support for democracy – 67% prefer democracy to any other form of governance (Afrobarometer, 2015) – democratic accountability is undermined by low levels of transparency, riven by vote-buying, and blighted by adverse selection into politics. For example, a USAID survey in 2015 indicates that 49% of citizens believe that “many” or “almost everyone” accepted gifts from parties in exchange for their vote and that 35% were personally given money in exchange for their vote. The focus in this paper is on the House of Representatives, where each of 73 electoral districts elects a single representative for a 6-year term in a first-past-the-post electoral system. Representatives are rewarded handsomely: one NGO estimates their annual salary at well over \$200,000 (Liberian Lawmakers Watch, 2017), and they indirectly gatekeep access to the rich natural resource wealth of the country by controlling access to concessionary agreements⁶.

It is therefore unsurprising that lots of people want to be politicians. In the 2017 House elections there are 984 candidates across 26 parties: “Rest assured that this is not a healthy expression of diverse opinions. Everyone wants a piece of the pie.” (Washington Post, 2017). The number of candidates per district varies between 3 and 28. Once in office there is substantial variation in performance, with incumbents attending as few as 42% of House sessions. The result is widespread dissatisfaction with politician performance (IREDD, 2017). 67% report trusting their Representative ‘Not at all’ or ‘Just a little’, 38% believe that Representatives ‘Never’ listen to what citizens say, and 68% either ‘Strongly disapprove’ or ‘Disapprove’ of their Representative’s performance (Afrobarometer, 2015). These high private returns suggest that politicians face incentives to restrict the amount of information they provide about themselves. For an illustration, the Daily Observer newspaper (one of the country’s most prominent) built a ‘promises tracker’ ahead of the 2017 election where candidates could specify their policy platforms to appeal to voters and commit to implementing specific projects. As of the time of writing, no incumbents have done this.

Combining high personal returns to public office with low oversight of their behavior induces an adverse selection of candidates into the political market (Brollo et al., 2013). Even aside from the incentives faced by politicians to limit access to information about their actions and avoid credibly committing themselves to programmatic policies, mechanisms to hold politicians to account for poor performance are scant – at least partially because of a paucity of credible information about politics in much of the country. Over half the population is illiterate, and most lack education past the primary level, which is compounded by the centralization of the country’s political apparatus in Monrovia. With most politicians ensconced in the capital, far from their citizens, the diffusion of information about their behavior is limited - more than one focus group reported only seeing their Representative once every six

⁶A major corruption scandal in 2016, for example, revealed that both the Speaker of the House of Representatives and the Chair of the incumbent Unity Party had conspired to illegally broker an iron ore concession in exchange for bribes (Global Witness, 2016).

years during campaigning season. Reflecting this fractured link, during campaigning season incumbents orchestrate the mass turnout-buying (Bowles, Larreguy and Liu, 2017), and trucking of voters from the capital to their districts. Campaigning season is marked by local rallies where candidates travel from community to community distributing food, clothing, and money to generate support, while making local non-credible policy promises to build schools, hospitals, roads and anything else that will get them votes. As such, candidate campaigns broadly lack policy platforms but rather focus on cheap talk about local development – the broadcasting of campaign messages across entire districts at once remains rare. Access to politicians, once in office, is rendered difficult by both poor infrastructural conditions and a limited, though developing, telecommunications network. The result is that citizens not only disapprove of their Representatives, but they have little idea what they do: nearly half of citizens incorrectly believe that they control the County Development Funds (Liberia Accountability and Voice Initiative, 2016).

The result is a low-quality, low-information equilibrium where the link between citizen and representative is broken. So, what can be done? We posit a central role for the media, which plays the intermediating role in the provision of political information. Internet penetration remains low outside of the capital and literacy rates limit newspaper readership. Radio stations are at the core of access to information: in a 2016 survey, 83% of respondents said that the main way they learned about the County Development Funds was through the radio (Liberia Accountability and Voice Initiative, 2016) versus 1.1% through the internet and 0.4% through newspapers. Radio ownership is high at 83%, and 62% report listening to news on the radio every day (Afrobarometer, 2015). Underlying this is a fragmented and highly unregulated radio market.

A survey of the country's radio landscape in 2016 located 105 stations, the majority of which are community-owned and run. For context, Kenya has a population 15 times as large as Liberia but far fewer radio stations (Internews Liberia, 2016). Many of these emerged in the immediate post-war years to preach peace and reconciliation, as well as to encourage the resettlement and reintegration of internally displaced peoples. This form of media, while it holds substantial potential for fixing the link between voter and representative, remains underutilized. The lack of regulation, sporadic access to electricity and scarce sources of commercial revenues for stations means that they frequently become the mouthpieces of particular political figures and local firms, and some incumbents actually own their own radio stations.

Very few radio stations employ professionally-trained journalists, and most journalistic skills are gained through short workshops run by NGOs. Stations employ on average 10.8 employees, varying between 2 and 30, but salaries are paid by few stations and staff retention is low. Audience surveys are scarce, and a lack of content for many stations drives a predominance of call-in radio programming: the modal form of radio show is a basic talk show where the presenter invites listeners to call in to discuss local issues. With no internet access at most stations and limited research capacities, this form of content is easy to produce but conveys little credible information. Community radio stations hold on average 1.94 hours of 'news programming' per day, but most of this is taken up by call-in segments (Internews Liberia, 2016).

Consequently, while radio is clearly the modal medium to access political information, the quality of programming is low and many Liberians distrust the information they get from the radio. 41% of Liberians believe that the political information they get from their favorite radio station is ‘Somewhat honest and accurate’ or ‘Not at all honest and accurate’ (USAID, 2015). Interventions designed to target the quality of political information through these community radio stations, therefore, hold real promise: the radio stations can see the consequences of broadcasting more politically-relevant information on audience listenerships, and potentially lead to the development of more professionalized models of content production. With a large latent audience for such material, the returns for both citizens and the radio stations themselves are potentially sizable.

3 Debates

Targeting these three key issues – adverse selection of politicians, limited citizen information, and an underdeveloped media capacity – Internews Liberia led a nationwide debates initiative in the run-up to the October 2017 elections for House of Representatives⁷. This was the first time that debates had been held across the entire country, and the model was designed to push back against Liberia’s clientelistic equilibrium and towards the beginnings of a programmatic one.

First, Internews engaged Liberian partner organizations to take charge of organizing the debates: the Press Union of Liberia (PUL), Liberia Media for Democratic Initiatives (LMDI) and the Center for Media Studies and Peacebuilding (CEMESP). Debate responsibility was divided among these partners geographically. Each district was assigned a local journalist who would moderate the debate, conduct research in the district, and publicize the debates. Due to the number of candidates in some districts, if a district had more than 8 candidates then 2 debates were held (typically in the morning and afternoon of the same day). In a few districts, 3 debates were held since there were more than 16 candidates.⁸ When a district had more than one debate, candidates were randomly assigned to debates within the district, a source of variation which we return to below. Specifically, the experience of inviting candidates to attend to the debates suggests that we might be able to exploit whether the debate to which a candidate was invited was held first, as well as whether the incumbent was also assigned to the debate, to construct an instrument for debate participation that varies across candidates within a district. We develop this further later. In total, 129 debates were held across all 73 districts – Figure 1 shows one in Monrovia. The first debates took place in mid-August, and ran through until mid-September.

Each debate had at least one radio partner present to broadcast (and later rebroadcast) its content, as detailed below. Election-related violence is a concern in Liberia, and so to minimize the risk of conflict, the in-person audience for the debates was kept small: each candidate was given 5 tickets to invite their team, and total audiences rarely exceeded 100 people.⁹ The structure of the debates was simple: up to

⁷Some debates were also held for the presidential race by other organizations which are not the focus here.

⁸While this assignment rule suggests a potential RD-style design with variation in the size of debates, preliminary evidence suggests compliance with the 8-candidate threshold was varied.

⁹In a pilot debate held in Lofa without limits on the in-person audience violence broke out when a candidate stacked the



Figure 1: District debate (Montserrado D3)

five questions were asked to all candidates in attendance. Every candidate was given an opportunity to respond to each question, with time limits on responses of 3 minutes. The order in which candidates responded to questions was randomly assigned by drawing straws at the start of each debate, a source of variation which we return to below. Specifically, we expect to be able to exploit whether a candidate was selected to speak earlier, and particularly so when his or her debate was held first, to construct an instrument for debate participation that varies across candidates within a district. Since citizens have a limited attention span, those candidates might benefit more from debate rebroadcasting.

In terms of debate questions, candidates were asked to outline their top policy priorities, including their key promise for what they would do if elected. Then, the first two questions in each debate were standardized by Internews and its implementing partners. The first question relates to the management of the County Development Fund (CDF)¹⁰, which is poorly managed with little input from citizens and oversight on how the funds are used. Candidates were asked how they would improve the management of the CDF to benefit the citizens in their community. Second, candidates were asked about how they would spend their Legislative Support Project (LSP) funds, instituted in 2015. Each House Representative is given discretionary LSP funding to spend as they like on development projects in their district, but

audience with supporters.

¹⁰The CDF was officially renamed the County and Social Development Fund (CSDF) in 2015 – however, the partner organizations and candidates were more familiar with it being called the County Development Fund and so we used this name both in the debates and the survey instruments.

the completion of projects is poor and citizen awareness about the existence of the LSP is limited. After these standardized questions, candidates were asked 2-3 questions based on research conducted by the moderator in the district about locally-relevant issues and how the candidates will address them.

4 Intervention

In partnership with Internews Liberia, we cross-randomize two elements of the debates initiative at the between-district level, beyond the within-district sources of quasi-random variation discussed above (the splitting of candidates across debates, size of debates, and ordering of candidates within a debate). First, we generate random variation in the attendance of political candidates across debates by varying whether debates are assigned a to receive more intensive effort in persuading political candidates to attend¹¹. Second, we generate random variation in the share of a given district which is likely to hear the debates at least once by varying the intensity of debate rebroadcasting. The interventions were designed to build off parts of the debates initiative without depriving candidates or voters from opportunities they would have received absent an evaluation of the intervention: rather than experimentally varying the extensive margin of exposure, both interventions were designed to ramp up the intensity of activities already planned to facilitate their evaluation. Doing otherwise would have risked real ethical concern.

4.1 Invitations

First, we randomize the intensity of debate invitations to candidates. Candidates in control districts are contacted by the relevant Liberian organizing partner, who invited them to attend and provided logistical information about the debate. In treatment districts, on top of contact by the partner organization we:

1. Sent emails to all candidates with working addresses 4 days before their debate. The 'official' email invitation included Internews/USAID branding, details on the purpose of the debate and information on who to speak with if they had any uncertainties. 71% of candidates had working addresses, with contact details provided by the NEC.
2. Phone calls to all candidates 2 days before the debate. These were mostly conducted by a high-profile Liberian radio journalist who is widely known and respected by Liberian politicians. In these calls, candidates were reminded why they should attend the debates and their concerns about any elements of debate organization were addressed. Calls were attempted 5 times in the case of non-response. Successful calls were made to roughly 75% of all candidates in treatment districts.
3. SMS reminders to all candidates on either the evening before, or the morning of, the debate with information on where to go. Every candidate had a listed phone number.

As such, the invitation intervention carries both behavioral elements (reminding candidates about the debates, since campaigning season is busy) and more persuasive ones (reducing fears about the bias

¹¹Randomizing invitation effort at the candidate-level, while cleaner experimentally, would have raised serious concerns in terms of ethics and fairness to candidates

of debate moderators, arguing why it is their democratic duty to participate in the debates). While this intervention is at the debate-level – randomizing at the individual-level carries the previously-discussed fairness and ethical concerns – we do not expect homogeneity in the response of candidates within a race to the treatment. In particular, we should expect that the treatment is more likely to affect the marginal decision of ‘top’ candidates. For these candidates there are both campaigning opportunity costs to debate participation as well as potential downside from performing poorly in the debate. For smaller candidates with more limited resources, these debates offer a much clearer proposition.

It is worth mentioning that other constraints affected debate participation aside from the demands of the campaign trail. First, Liberia’s rainy season renders much of the interior districts near-impassable. For the many candidates who live in the capital, the difficulties of travel were frequently cited as the reason for non-attendance. This was compounded by the weak organizational capacity of some of the partner organizations implementing the debates, which resulted in frequent short-notice changes to debate logistics.¹² Further, during some of the later debates it became apparent that some control districts were *de facto* treated: the intervention was sufficiently easy to implement that evidence of its efficacy led debate partners and, sometimes, the coordinating NGO to imitate it in control districts.

Qualitative evidence from the phone calls to invite candidates to participate in the debates suggested that two sources of additional random variation contributed to the attendance decision: first, whether candidates had been assigned to the incumbent’s debate. Given the allocation of candidates across debates in a district, it is effectively a random variable whether a given candidate was assigned to debate the incumbent in those cases where more than one debate was held. Many candidates asked whether the incumbent had been assigned to their debate, specifically seeking to debate against them – attacking the incumbent could offer both credibility by appearing on the same stage and, in some sense, makes the debate ‘easier’ by creating a focal point for easy attacks. Second, several candidates expressed a preference for participating in the first debate in their district, perceiving that the first debate might garner the most attention. We return to these below. While, as mentioned above, the intervention was intended to generate variation in debate participation at the district level, we anticipate that this quasi-random variation on whether the candidate was assigned to the first debate, or together with the incumbent, might yield variation across candidates within a district.

4.2 Rebroadcasting

Second, we randomize the intensity with which the debates are broadcast. In control districts, debates are broadcast live by the partner community radio station and at most 2 times in the two weeks following the debate. In treatment districts, debates are broadcast an additional 10 times in a staggered fashion leading up to the election with 5 rebroadcasts in the 10 days before the election. Due to a lack of regulation and consolidation, there are a surprising number of community radio stations in Liberia – well over 100

¹²As one candidate wrote in reply to the official email invitation, “I acknowledge receipt of your invitation to attend the debate in District 2 initially slated for the 5th, then the 20th and now the 18th of September. I don’t know why the changes without considering the schedules of the participants who are the subject of this discourse.”

in a country of 4 million. We selected 43 out of this set to rebroadcast the debates, selecting based on the signal strength of the station to maximize audience sizes and discounting any stations which were owned by candidates for office in that district.¹³

Contractual agreements were made with these 59 stations to ensure that they rebroadcast the debates in their entirety without selectively editing (except to reduce dead air time between questions) and stations were provided high-quality recordings of each debate in case their recording equipment had failed during the debate. A youth network, Naymote, was given the rebroadcasting schedule and mobilized individuals in every district to tune in at the stated time to ensure that the debates were broadcast as agreed. Reports from these monitors suggest that around 90% of the scheduled rebroadcasting in intensively-rebroadcast districts took place on time. The specific times and dates of the rebroadcasts in intensive districts were selected to maximize audiences by choosing ‘shoulder’ and ‘peak’ times for Liberian radio listenership: between 6am-10am in the morning and 6pm-10pm in the evening. We varied the time of day across rebroadcasts within these high-listenership bands to further maximize the intensity of the debate broadcasting.

4.3 Treatment assignment

To assign treatment conditions, first we pre-stratify based on the debate organizer (PUL, LMDI, CEMESP or LMD). This is because the capacity of the debate organizers varied substantially in terms of their ability to attract candidates, organize the logistics of the debates and in the quality of moderation. Second, we block on a set of pre-treatment covariates at the district-level.¹⁴ In Table 4.3 we report the pre-treatment balance across covariates, where regressions are run of the form:

$$y_d = \beta_1 T^I + \beta_2 T^R + \eta_b + \epsilon_d \quad (1)$$

Where y_i is the pre-treatment covariate y in district d , T^I is an indicator for the candidates in the district receiving the invitation intervention, T^R is an indicator for the district being assigned to the rebroadcasting intervention, η_b are block fixed effects, and ϵ_d are heteroskedasticity-robust standard errors. The coefficients β_1 and β_2 are reported in the balance table. In the treatment assignment selected, 1 (2) out of 47 covariate(s) are imbalanced at the 5% (10%) level between treatment and control for the invitation intervention (β_1). 1 (3) out of 47 covariate(s) are imbalanced at the 5% (10%) level for the rebroadcasting intervention (β_2). Overall, the few imbalances are consistent with chance.

We also report balance tests based on regressions of the following form, where we use three indicators that capture treatment assignment to invitations and rebroadcasting: (High, Low), (Low, High),

¹³Very few stations were discounted in this way, primarily because they could not be guaranteed to replay the debates in full with no editing. In the most rural areas, the partner NGO even had to bring 2 radio stations back on air by repairing signal transmission equipment which had been recently damaged by lightning.

¹⁴We use the R package “blockTools” to assign blocks, by debate partner, based on Mahalanobis distance. We block on the initially-planned week of the debate, number of candidates, whether the incumbent is seeking re-election, log of registered voters in that district, number of debates to be held in district, vote share for top 3 candidates in 2011, vote share HHI in 2011, turnout in 2011, share of candidates who ran in 2011, log population density, GSM phone coverage, share of citizens who own a radio, and share of citizens who frequently get news from the radio.

(High, High) such that (Low, Low) is the omitted category:

$$y_d = \beta_1(H, L) + \beta_2(L, H) + \beta_3(H, H) + \eta_b + \epsilon_d \quad (2)$$

Table 4.3 reports the coefficients on β_1 , β_2 and β_3 . η_b are block fixed effects, and ϵ_d are heteroskedasticity-robust standard errors. 2 (4) covariates out of 47 are imbalanced at the 5% (10%) level between the (High, Low) group and (Low, Low) (β_1). 3 (6) covariates are imbalanced at the 5% (10%) between the (Low, High) group and (Low, Low) (β_2). And 1 (3) covariate(s) are imbalanced at the 5% (10%) level between the (High, High) group and (Low, Low) group (β_3). Overall, the few imbalances are consistent with chance.

5 Data

We are conducting a panel survey of around 4,000 citizens across all 73 electoral districts. Due to time constraints and the difficulty of traveling within the country during the rainy season, we opted for phone-based surveys using Computer-Assisted Telephone Interviews (CATI). Our survey implementing firm, GeoPoll, samples phone numbers from the universe of active cell phone numbers for the country's largest mobile network (Lonestar/MTN). With a team of university-educated enumerators in Monrovia, phone interviews are conducted in simple English and take around 20 minutes for the baseline and 30 minutes for the endline. The distribution of observations per electoral district naturally reflects cell phone penetration and rurality, but every district has a targeted minimum of 30 observations. Random samples pre-stratified by county are released to the enumerators, who try each number 5 times to attempt to complete interviews. If individuals are not free at the time of the call, they are called back at a later time. Respondents are given a token of appreciation of \$0.50 in phone credit in exchange for their time. Survey eligibility is based on providing evidence that the respondent is a registered voter. Preliminary estimates from baseline data suggest survey completion rates of 20-25% out of the total number of calls made. We interview around 5,000 respondents in the baseline to account for expected attrition.

In Figure 2 we show a timeline of the debates initiative and data collection. The number of debates on different days is shown on the y-axis, varying between 0 and 9 on a particular day. Data collection began in early August, right before the first debates. The bulk of data collection was completed by early September. However, baseline data collection continued through to the end of September to target the final electoral districts which proved hard to sample. Since the sample was stratified at the county-level, sampling within particular districts proved difficult especially when one county contained both urban and rural districts: in these cases most calls went to those in the urban districts, and so achieving sufficient sample in the more rural districts took longer than anticipated. We do not consider the overlap of the baseline survey and the live debates themselves to be a major concern. First, we can control for baseline debate exposure using the date on which respondents were interviewed. Second, this generates within-district variation in exposure to debates, which we describe more below, and lets us get at its

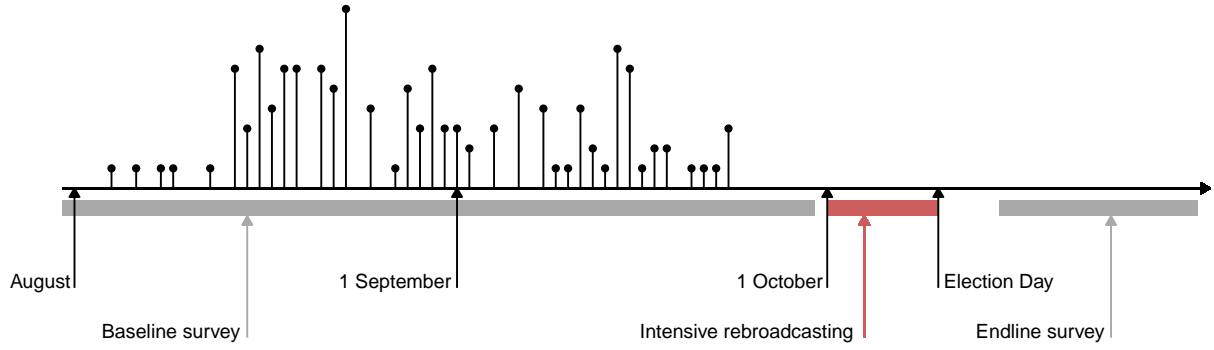


Figure 2: Timeline of intervention (*y-axis plots number of debates on a given day*)

short-term effect. And third, the intensive rebroadcasting of debates took place in October, by which time essentially all data collection was completed. We anticipate that the bulk of respondents actually heard the debates during this period.

While the survey data will be used for most of our key analysis, several other data sources will be used as detailed below. We refer to particular baseline and endline survey variables using B/E followed by the question number (e.g. B-Q6, E-Q30). We group data primarily according to the family of variables defined in the baseline and endline surveys. For each family of variables, we will use both a composite z-score as is standard practice, along with individual variables. When we test individual variables, then we will use standard adjustments for multiple comparisons.

5.1 Intervention compliance

For our *de facto* first stages, we measure whether the twin interventions actually worked. For candidate invitations, we simply collect data on which candidates turned up to the debates, based on narrative debate reports as well as full transcripts of each debate. For radio rebroadcasting, we measure – based on NAYMOTE reports – how many times each of the debates was broadcasted on partner radio stations. The number of radio broadcasts will then be instrumented with treatment assignment, as detailed below.¹⁵ We focus on the number of broadcasts as the measure of broadcast intensity rather than a measure of whether respondents have heard the debate for two reasons. First, because listening to the debate is also a function of which candidates attend. Second, because the policy-relevant parameter is how intensive radio coverage affects political outcomes – *not* just how an individual hearing a debate affects attitudes.

¹⁵We also intend to ask rebroadcasting stations whether they created any content based on the debates – e.g. discussing them in a news segment. We also intend to survey non-partner radios to establish whether they broadcast any on the Internews debates or their content some other way. While we do not anticipate that such was the case since the each of the debates was only made to the partnering radio hired to broadcast it, there is the remote possibility that they recorded the debates themselves and subsequently rebroadcast them or some of their content.

5.2 Candidates

In the panel survey, we ask respondents about specific candidates in their district and the unit of analysis, detailed below, will generally be the respondent-candidate dyad. This is both because asking about up to 25 candidates would make the survey prohibitively time-consuming, and because we have a theoretical reason to expect that the invitation intervention should differentially affect the attendance decision of top candidates. So, for each candidate running, we construct an indicator variable for whether the candidate is predicted to place in the top 3 in that district. We construct this indicator as follows, in a sequential fashion until there are 3 per district:

1. If a candidate is the incumbent.
2. If the candidate ran in the 2011 House election and placed either 2nd or 3rd.
3. If the candidate is from a top party: sequentially the incumbent Unity Party (UP), Coalition for Democratic Change (CDC), Liberty Party (LP), the Alternative National Congress (ANC) and the All Liberia Party (ALP).

This process resulted in 3 selected candidates in all districts. For data on the policy promises of candidates, we code up which candidates attended out of those invited, the order in which candidates were asked questions, and information on the in-person audience. We also have data on all the debate locations, present media partners, and the demography of the in-person audience. In partnership with the Daily Observer newspaper, we are having transcripts from every debate transcribed by trained journalists – the newspaper is using these to extract promises from each candidate which are likely to be used in a follow-up study. As such, we will use these transcripts to code up the top policy priorities of each candidate, their policy prescription to improve the management of the Community Development Fund, and their intended use of the Legislative Support Projects funds. To supplement this data, we intend to run a candidate survey shortly after the election to solicit the top priorities of non-participating candidates. Aside from these priorities, we intend to collect information on the educational attainment of different candidates, their professional experience, their wealth, and their experience in politics.

5.3 Debate exposure

We assess the exposure of respondents to the debates by considering variables which directly reflect listening to the debate. This includes a binary variable for whether respondents have heard the debate between candidates on the radio (B-Q7, E-Q14), how often the respondents heard the debate (E-Q15), accurately knowing how many candidates participated (E-Q16), whether they report changing their mind based on the debate (E-Q18). About particular candidates, we ask whether they heard the 3 selected candidates participated in the debate to verify listening (E-Q39.2, E-Q39.4, E-Q39.6), whether respondents know anything about their policy promises (B-Q15, B-Q17, B-Q19, E-Q29, E-Q31, E-Q33), whether they responded that they knew something about each candidate's competence (B-Q21, B-Q23, B-Q25, E-Q33,

E-Q35, E-Q37), and whether respondents know the name of their incumbent (B-Q12). Lastly, we also measure whether respondents know more about the management of the Community Development Fund (CDF), which was discussed in the debates: who manages it and the legal requirements for consulting citizens and reporting on expenditure (B-Q9, B-Q10, B-Q11, E-Q9, E-Q10, E-Q11).

5.4 Knowledge about candidates

We use a family of variables to assess whether respondents learned about the policy promises and competence of candidates. First, we code whether respondents are more likely to correctly name the top policy priorities of participating candidates in endline compared to baseline (B-Q15, B-Q17, B-Q19, E-Q29, E-Q31, E-Q33). Second, we look at whether respondents have more precise beliefs about candidates: both whether respondents can assess and are more sure of the policy promises of candidates (B-Q16, B-Q18, B-Q20, E-Q28, E-Q30, E-Q32), and whether they can assess and are more certain about the competence of each candidate (B-Q22, B-Q24, B-Q26, E-Q34, E-Q36, E-Q38).

5.5 Beliefs over candidate competence

Aside from knowing whether respondents are more knowledgeable and certain about the competence of candidates, we seek to measure how beliefs over candidate competence were affected by the debate. For this, we use the prior and posterior beliefs about the competence of candidates (B-Q21, B-Q23, B-Q25, E-Q33, E-Q35, E-Q37). With this, we use a measure of debate performance in the interactive specifications outlined below. To measure debate performance, we ask respondents who they believe won the debate at endline (E-Q17) and employ this in a jack-knife estimator as detailed below, where the relative share of respondents stating that a particular candidate won the debate measures performance. Second, we intend to code the quality of debate participation by evaluating, e.g. the precision of their responses to questions based on transcripts of all debates – however, using an expert panel as in other debate-based evaluations such as Bidwell, Casey and Glennerster (2016) and Izama and Raffler (2016) is unfeasible here due to the number of debates actually held. We also intend to construct a measure of candidate “quality” for the incumbent, this will be drawn from IREDD legislative report cards measuring performance in office and for other candidates will be drawn from the candidate survey.

5.6 Voting behavior

Next, we evaluate any impacts on voting behavior. First, we assess the extent of the preference alignment between citizens and candidates by asking respondents to name what they consider to be the top issues in their districts (B-Q15, E-Q26), and constructing a variables indicating whether particular candidates also focused on these issues – a binary indicator for whether top issues are shared, and continuous measures of how closely preferences are aligned. We also measure preference alignment in two other ways: we ask respondents what they believe the biggest issue with the management of the CDF is (B-Q14), which we

match to candidate promises over how to fix the CDF, and use candidate responses to how they intend to spend the LSP funds matched to respondent's top policy priority areas. We use this 'preference match' variable to assess whether respondents are more likely to shift towards candidates with whom they share policy concerns. For this, we measure baseline candidate vote selection (B-Q27), and in the endline ask who they actually voted for (E-Q45). Aside from preference alignment, we use the same measures of debate performance as above. Second, we are interested in whether the types of justification cited by respondents are affected by treatment, and so ask at both baseline and endline the three top reasons for their vote selection (B-Q29.1, B-Q29.2, B-Q29.3), specifically focusing on whether respondents cite the campaign promises of candidates in explaining their vote choice. To verify voting by respondents, we ask a series of verification questions (E-Q43, E-Q44.1, E-Q44.2, E-Q44.3). Lastly, for voting behavior, we will also use geolocated polling place-level electoral returns with vote shares/turnout for each candidate.

5.7 Voter coordination

To evaluate outcomes around voter coordination, we ask how often respondents discussed political issues with friends (B-Q8, E-Q8). For more targeted information on the debates, we collect data on how often respondents discussed the debates with acquaintances (E-Q19), whether this led them to agree on a particular candidate (E-Q20), and whether this led them to change their mind about voting (E-Q18).

5.8 Campaigning

We ask respondents about their experiences of campaigning in the election. We ask how often respondents heard advertising about particular candidates on the radio in the two weeks before the election (E-Q39.1, E-Q39.3, E-Q39.5), and whether agents from the candidate visited their community and distributed flyers (E-Q41.1, E-Q41.2, E-Q41.3). Respondents also express how often they believe that others in their community sold their vote (E-Q40.1, E-Q40.2, E-Q40.3).

5.9 Media consumption, attitudes, and electoral institutions

To measure media consumption and attitudes towards the media, we ask citizens for how often they listen to the radio (B-Q6, E-Q6), how often they sought information from non-radio sources (E-Q7), which stations they listen to and when (E-Q5). To measure attitudes towards the media sector, we evaluate whether respondents believe that the media improves the competence and representation of candidates (E-Q21, E-Q22), whether the media was biased in its coverage of different candidates (E-Q23, E-Q24), and the extent to which respondents trust the information they get from radio stations (E-Q25). We additionally ask whether the respondent believe that elections improve the quality of elected officials (E-Q12, E-Q13).

5.10 Other respondent covariates

For each respondent, we ask for their Voter Registration Center (VRC) number reflecting their polling place on election day. Combined with geolocations of each polling place, geolocations of all radio stations, and topographical data, using radio propagation software,¹⁶ we can verify whether a given respondent/VRC received signal from the partner radio station. This will also allow us to evaluate the rurality of different respondents. We ask respondents what their top three radio stations are, to verify whether they are likely to have heard the debates, and what times of the day they listen to the radio (E-Q5). Combined with information on when the debates were actually rebroadcast, we can use this to generate within-district variation in exposure. Data is also collected on basic socio-demographics of respondents (B-Q1, B-Q2, B-Q3, B-Q4, E-Q2, E-Q3, E-Q4).

5.11 Radio stations

Aside from data on the radio stations used to calculate coverage of respondents, we collect a series of covariates about the radio stations themselves: their ownership type, the amount of news programming they broadcast, their age, signal strength, how many employees they have. This data is drawn from internal Internews documentation. We also intend to implement a brief survey of radio stations in cases where this information is missing, to assess whether contracted stations broadcast any other information related to the debates in the run-up to the election, whether non-contracted stations broadcast anything relating to the debates (which we consider unlikely), and information on their political affiliations, if any.

6 Estimation

There are two potential basic estimation approaches for analyzing the survey data. In the first (equation 4), we regress outcomes onto two treatment variables for whether the unit of observation was in a district assigned to intensive invitations, T^I , and whether the district was assigned to intensive rebroadcasting, T^R .¹⁷ In the second (equation 5), we use three indicators that capture treatment assignment to invitations and rebroadcasting: (High, Low), (Low, High), (High, High) such that (Low, Low) is the omitted category.¹⁸ Taking the case where the respondent-candidate is the unit of observation, we have that:

$$y_{icd} = \beta_1 T^I + \beta_2 T^R + \eta_b + \theta_e + \epsilon_{icd}, \quad (4)$$

$$y_{icd} = \beta_1(H, L) + \beta_2(L, H) + \beta_3(H, H) + \eta_b + \theta_e + \epsilon_{icd}, \quad (5)$$

¹⁶CloudRF was used for initial signal propagation mapping.

¹⁷We also anticipate constructing a continuous version of T^I by leveraging the other sources of random variation in the attendance of candidates: assignment to the incumbent's debate, and assignment to the first debate in the district. We will create a weighted index based on these random variables, which also can also be applied to construct a continuous version of equation 5 below.

¹⁸A third possibility (equation 3) instead considers the interaction of the marginal treatments ($T^I \times T^R$):

$$y_{icd} = \beta_1 T^I + \beta_2 T^R + \beta_3(T^I \times T^R) + \eta_b + \theta_e + \epsilon_{icd}, \quad (3)$$

However, such a possibility is dominated by the specification in equation 5 from a statistical-power standpoint.

where y_{icd} is the outcome for respondent i regarding candidate c in district d , η_b are randomization block fixed effects and θ_e are enumerator fixed effects.¹⁹ Standard errors are clustered at the electoral district level throughout. We consider two approaches to weight observations: either by the number of registered voters in the district over the number of surveyed registered voters, or by the inverse of the number of surveyed registered voters.

These estimation approaches extend to cases where the respondent is the unit of observation, y_{id} , and where the candidate is the unit of observation. For the rest of the PAP we use equation 5 as the base specification but modifications can just as easily be applied to other estimation approaches, as well as the IV approach below. We generally make hypotheses regarding β_3 since we expect effects to be strongest here, but should also expect coefficients on β_1 and β_2 to be in the same direction.

While equations 4 and 5 are all effectively reduced-form, we also have the option of running IV regressions where our endogenous variables are the attendance of candidate c in district d , \widehat{Attend}_{cd} , and the number of broadcasts of district d 's debates, $\widehat{Broadcasts}_d$, which we instrument using the relevant treatment assignments such that equation 5 is the first stage for the following estimation:

$$y_{icd} = \beta_1 \widehat{Attend}_{cd} + \beta_2 \widehat{Broadcasts}_d + \beta_3 (\widehat{Attend}_{cd} \times \widehat{Broadcasts}_d) + \eta_b + \theta_e + \epsilon_{icd}, \quad (6)$$

We also make use of specifications where we interact treatment assignment with particular covariates X_{cd} that vary within-district, for example at the candidate level c in district d :

$$y_{icd} = \beta_1 ((H, L) \times X_{cd}) + \beta_2 ((L, H) \times X_{cd}) + \beta_3 ((H, H) \times X_{cd}) + \gamma_1 X_{cd} + \mu_d + \eta_b + \theta_e + \epsilon_{icd}, \quad (7)$$

where the main effects on treatment indicators drop out due to the inclusion of district fixed effects μ_d . In some cases, the variation we seek may stem from cross-district variation, in which case we drop the district fixed effects, and thus also estimate the main effects on treatment indicators. Similarly, for a few cases, we have covariates X_d , which only vary at the district level, in which case we are forced to drop the district fixed effects.

Also, for many survey outcomes, we have panel responses pre- and post-intervention, and we use Δy_{icd} as an outcome instead, but the estimating equations remain the same.

As we discussed above and Figure 2 shows, the baseline data collection took place while the debates and initial rebroadcasting were conducted, but before the intensive rebroadcasting of debates. This timing generates within-district variation in exposure to debates and initial rebroadcasting, which we can exploit to estimate their short-term effect. Specifically, we use the following local-regression discontinuity design:

¹⁹We also anticipate running versions of these estimating equations where we control for the date of the baseline interview: $\mathbb{1}(b.debate)_{icd}$ is an indicator variable for whether respondent i was interviewed in the baseline after the debate had actually been held which was the case in several districts. We expect that the large majority of respondents heard the debates during the intensive rebroadcasting period in October, but will also include this linearly as a control variable for both robustness and power reasons. Additionally, we intend to estimate specifications including district-level covariates. We anticipate limited gains from including covariates due to our inclusion of block fixed effects, but will nonetheless assess whether doing so improves the precision of our estimates.

$$y_{icd} = \beta_1 D_{id} + \beta_2 distance_{id} + \beta_3 D_{id} \times distance_{id} + \mu_d + \eta_b + \theta_e + \epsilon_{icd}, \quad (8)$$

where D_{id} indicates whether the debate was conducted and rebroadcast by the time respondent i was surveyed, and $distance_{id}$ is the distance (in days) from the time when respondent i was surveyed and the time the debate was conducted and rebroadcast. $distance_{id}$ takes negative values if the debate was not yet conducted and positive otherwise. Given the need to use a narrow bandwidth (in terms of days) to convincingly estimate the treatment effect at the threshold, if this leaves us with very few observations, we will also use a specification excluding the linear control for $distance_{id}$, and its interaction with D_{id} . Either way, the coefficient of interest is β_1 .

Last, note that, since all our hypotheses below are one-sided, we consequently consider one-tailed tests to assess their empirical relevance.

7 Hypotheses

7.1 Debate exposure and knowledge about candidates

First, we seek to evaluate whether random variation in encouragement to participate in debates, combined with random variation in the broadcasting of debates, affects actual participation in debates and the actual broadcasting of our debates. There are two variations of y_{icd} , the respondent-candidate outcome, we intend to use to maximize power given the limited sample size of districts. In the baseline specification as above, we treat the respondent-candidate as the unit of observation. In a secondary specification, we exclude responses about the incumbent (if they are seeking re-election) in the respondent-candidate specifications. The idea is that citizens are much more likely to already know about the incumbent compared to their challengers and so the effect of treatment assignment is likely to be more muted about those candidates. For these outcomes, we use index variables described in section 5.3 and 5.4.

I1. Citizens were more intensively exposed to the debates in more intensively treated districts: $\beta_3 > 0$ in equation 5.

We measure the intensity of debate exposure using variables described in the Data section: whether individuals have heard the debates, how often they heard the debates, and how accurately they knew the number of candidates who participated. We anticipate this hypothesis as a basic sense-check that the intervention had some effect: respondents in districts where the debate was broadcast just live and shortly afterward, well before the election, should have had a much lower chance of exposure to the debates. We estimate this regression at the respondent-level y_{id} .

I2. Citizens know more about policy in more intensively treated districts: $\beta_3 > 0$ in equation 5.

Again with the same estimation, we hypothesize that individuals in districts more intensively treated are likely to be exposed to information about policy areas – specifically the management of the CDF – which came up in every debate. We measure this using the family of variables regarding the CDF outlined above which are factual questions about the management of the CDF.

I3. Citizens know more about candidates in more intensively treated districts: $\beta_3 > 0$ in equation 5.

Individuals more intensively treated with the debates intervention may be expected to learn more about particular candidates running for office. We use respondent-candidate dyads and use whether respondents heard specific candidates in the debates, how often they heard about that candidate on the radio in the 2 weeks before the election, whether respondents improved in their knowledge of candidate policy promises (using Δy_{icd} as the outcome variable), whether more knew the name of the incumbent, and whether they were less likely to respond “do not know” (DNK) to questions about candidate competence. We also employ data on the precision of respondent beliefs over candidate policy promises and competence, as outlined in section 5.4.

Fourth, we use within-district variation stemming from the overlap of the baseline survey with the debates to provide extensive-margin evidence on the effects of debate exposure on political information about candidates and policy. In particular, we subset just to the baseline data and conduct the local regression discontinuity design in equation 8. With this within-district variation in exposure, we expect that relevant political information outcomes should be positively affected by being interviewed after the debate itself happens.

I4. Citizens interviewed after the debate were more likely to have heard the debate, to know more about policy, and to know more about candidates: $\beta_1 > 0$ in equation 8.

We expect that the β_1 should be positive since this represents the treatment effect of being interviewed just after, rather than just before, the initial live broadcasting of the relevant debate. We expect this treatment effect to be positive but it will likely be substantively small: the share of respondents who are likely to hear the debate live is small, and we expect most respondents to hear the debates in the intensive-rebroadcasting period.

7.2 Beliefs about candidates

Next, we evaluate whether the intervention affected how beliefs about the competence of candidates were affected by the debates, as well as how respondents made their vote choice. First, we evaluate how respondents update about the competence of candidates based on debate performance conditional on debate participation. The simplest form of these regressions is based on equation 7, where $X_{cd} = \text{Performance}_{cd}$ is a measure of debate performance for a particular candidate and we use Δy_{icd} as the outcome variable, since all these outcomes are measured in both baseline and endline.

B1. Citizen evaluations of candidate competence are increasing in the debate performance of participating candidates: $\beta_3 > 0$ in equation 7.

Where we construct the variable Performance_{cd} as a continuous variable measuring the proportion of respondents in that district (apart from respondent i) who cited candidate c in district d as the ‘winner’ of the debate, based on the endline survey. Given the number of debates involved, we consider using an expert panel to be infeasible here as has been done in some previous studies. However, as a secondary check, we will code, e.g., the specificity/quality of the policy promises made across candidates and

debates. Evaluations of candidate competence are measured using the variables outlined in section 5.5. We should note the possibility that debate performance is endogenous to the invitation intervention – debate performance is endogenous to who shows up to the debate. First, we can assess whether measures of candidate quality (described below) are balanced across treatment groups. If they are balanced, such that the invitation affected participation decisions more so of top candidates, but not differentially across quality among this group, then this is less of a concern. If we remain concerned about selection effects, then we have several instruments available: the assignment of candidates to incumbent debate, the first debate in the district, and the number of candidates in the corresponding debate while controlling flexibly for the number of candidates in the race—it should capture plausibly-exogenous variation in the number of candidates in each debate since the maximum number of candidates invited to each debate was arbitrarily limited.

B2. Citizen evaluations of candidate competence are increasing in candidate quality in intensively treated districts: $\beta_3 > 0$ in equation 7.

In this specification, we instead set $X_{cd} = \text{Quality}_{cd}$ and construct a measure of candidate qualification/education based on a candidate survey. This ought to serve primarily as a sense-check for B1, since we might expect that measures of candidate quality will correlate strongly with debate performance. Next, we consider specific consequences for the incumbent where there is no variation within X in a given district, so now our interaction uses X_d :

B3. Citizens evaluations of incumbent competence are increasing in debate performance in intensively treated districts: $\beta_3 > 0$ in equation 7.

B4. Citizens evaluations of incumbent competence are increasing in incumbent quality in intensively treated districts: $\beta_3 > 0$ in equation 7.

These specifications require subsetting down to responses about the incumbent (running in 66/73 races), and hence district fixed effects will also drop from these specifications. We have reason to believe that incumbents attending the debate could damage their electoral prospects conditional on them being low-quality or poorly performing. The debate performance of the incumbent will be coded as above, while the measure of their quality will be based on legislative report cards indicating the frequency of their attendance of House sessions.

We consider within-district variation in whether individuals were interviewed before or after the debate took place at baseline. We use equation 7 and set $X = \mathbb{1}(b.\text{debate})_{icd}$ with Δy_{icd} as the outcome variable: we should expect that changes between baseline and endline are smaller for those who were interviewed at baseline after the first debate had taken place.

B5. All effects are smaller among those respondents who were interviewed at baseline after the first debate had taken place: $\beta_3 < 0$ in equation 3.

We use the outcome variables listed in B1-B4 to assess this outcome.

Finally, we also use within-district variation stemming from the overlap of the baseline survey with the debates to provide extensive-margin evidence on the effects of debate exposure on citizen evaluations

of candidate competence.

B6. All effects are larger when citizens interviewed after the debate: $\beta_1 > 0$ in equation 8.

We also use the outcome variables listed in B1-B4 to assess this outcome.

7.3 Preferences and voting behavior

Next, we consider whether citizens actually changed their vote choice for particular candidates at all. At the respondent-level, using the same specification from the previous section, where we set $X_{cd} = \text{Performance}_{cd}$ and use the respondent-candidate unit of observation, with Δy_{icd} as the outcome variable:

V1. Citizens are more likely to change their vote towards a particular candidate when they perform well in the debate: $\beta_3 > 0$ in equation 7.

This is a simple extrapolation of previous hypotheses: if think that the intervention affected perceptions of candidate competence, then we should expect that those respondents who change their vote choice are likely to do so towards candidates who performed better in the debate. In the following specification we set $X_{icd} = \text{Match}_{icd}$, i.e. the interaction is now with the extent of preference alignment between respondent i and candidate c .

V2. Citizens are more likely to change their vote towards a particular candidate when they learn their preferences align with the policy platform of the candidate: $\beta_3 > 0$ in equation 7.

Here the outcome variable is a binary variable for the respondent's change in vote choice intention for candidate c . The Match_{icd} variable takes several forms as explained in 5.6.: based on alignment surrounding top priorities, LSP funding, and ways to improve the management of the CDF.

Second, we evaluate how the intervention affects the reasons cited for particular vote choices. We run regressions of the form identical to equation 5, where y_{id} is based on the reasons cited for particular vote choices. In particular, we construct an indicator variable for whether the respondent cites the *campaign promises* of a particular candidate as justifying their vote choice, as outlined in section 5.6.

V3. Citizens are more likely to cite campaign promises as their reason for voting for their chosen candidate in intensively treated districts: $\beta_3 > 0$ in equation 5.

The intuition here is simple – in districts where debate broadcasting represents a greater share of airtime, citizens may substitute towards considering these campaign promises to be a more important factor in determining their vote choice. As in the previous section, we exploit the within-district variation in whether individuals were interviewed before or after the debate took place at baseline. Again use equation 7 and we set $X = \mathbb{1}(b.\text{debate})_{icd}$ with Δy_{icd} as the outcome variable: the intuition is as before and we expect less updating for those interviewed later in the baseline.

V4. All effects are smaller among those respondents who were interviewed at baseline after the first debate had taken place: $\beta_3 < 0$ in equation 3.

We use the outcome variables listed in V1-V3 to assess this outcome. We also evaluate outcomes at the candidate-level on voting outcomes. This shifts our unit of observation to candidate c in polling

place p and where data comes from polling place-level electoral returns, but the basic estimation remains as in equation 7. Now, however, the heterogeneous X_d regarding debate performance is given by the share of respondents in district d who stated c performed best in the debate.

V4. Candidates who relatively better receive more votes in intensively treated districts: $\beta_3 > 0$ in equation 7.

Similarly, we can define a district-level $Match_{cd}$ variable for the relative share of priority areas cited by respondents which are shared by candidates.

V5. Candidates whose policy promises match the preferences of voters receive more votes in districts where more citizens hear the debates: $\beta_3 > 0$ in equation 7.

Both V5 and V6 are simple translations of the individual-level hypotheses.

7.4 Campaigning

Next, we consider whether candidates have an endogenous campaigning response to debate participation. We expect to find null effects on campaigning efforts, focused on vote-buying, by candidates. While the Bidwell et al. (2017) study finds strong effects on campaigning, we conjecture that by disseminating the debate through community radio stations we minimize the capacity of candidates to target campaigning at particular villages which are more likely to have heard and discussed the debates. For this, we estimate respondent-candidate dyadic regressions of the form in equation 4.

C1. Citizens exposed to a higher intensity of debates do not experience more intensive campaigning by candidates in the run-up to the election: $\beta_3 = 0$ in equation 5.

We measure the outcome, campaigning effort experienced, by using variables described in section 5.8, which measure how often other individuals in the respondent's community sold their code to candidate c and how often they saw representatives from candidate c in their local community. It may even be the case that debate participation actually reduces the ability of candidates to buy votes from citizens. First, there are time-based opportunity costs to debate participation on the ability of candidates to travel to different towns in their district at the height of the campaigning season. While these are likely to be small, candidates who participate in debates may also commit themselves – at some level – to campaigning on more programmatic basis rather than entirely focusing on the distribution of cash and policy promises which vary by village. This would imply $\beta_3 < 0$, but we expect the overall effect to be minimal if at all present.

7.5 Media consumption, attitudes, and institutions

Next, we consider how the intervention, by changing the composition of media consumption, affected deliberation leading up to the election – and how the nature of the programming affected attitudes towards the media more broadly. We estimate regressions of the form in 5 to evaluate whether the rebroadcasting treatment assignment affected how often respondents listen to the radio, how often they

sought information from non-radio media sources about candidates, and how often they discussed the debates with others.

M1. Individuals in intensively-rebroadcast districts increased their consumption of media about House of Representatives candidates: $\beta_3 > 0$ in 5.

Using Δy_{icd} or Δy_{id} as the outcome variable, we will use binary variables for whether respondents got political information from the radio every day, whether they got information from non-radio sources every day, and how often they discussed politics with their family.

Second, we investigate whether exposure to the radio debates affected political behavior through a coordination channel: first – whether respondents are more likely to have discussed the debate and coordinated their vote choice when the rebroadcasting was more intensive, with estimations just as in equation 5.

M2. Individuals are more likely to have discussed the debate with others in intensively-rebroadcast districts: $\beta_3 > 0$ in equation 5.

M3. Individuals are more likely to express changing their mind based on the debate in intensively-rebroadcast districts: $\beta_3 > 0$ in equation 5.

M4. Individuals are more likely to have coordinated their vote choice based on the debate with others in intensively-rebroadcast districts: $\beta_3 > 0$ in equation 5.

These outcome variables come from the endline survey as described in Section 5.9. We will verify this outcome through the endline question asking how many times the respondent heard the debate on the radio.

Third, we consider whether the intervention changed the perception of citizens about the media itself, given the content represents a substantial shock to the dominance of biased talk shows on the airwaves. We hypothesize that individuals in districts where debate broadcasting was intensive in the weeks before the election have more positive attitudes about the contribution of the media to electoral credibility and democracy more generally. Regressions take the form of equation 5.

M5. Individuals believe that the local media is less biased, more trustworthy and provides more equal coverage of political candidates in intensively-rebroadcast districts: $\beta_3 > 0$ in equation 5.

M6. Individuals believe that the local media helps citizens choose more effective representatives in intensively-rebroadcast districts: $\beta_3 > 0$ in equation 5.

Outcome variables for both M5 and M6 come from section 5.9. We expect that the shock to the nature of radio programming at a salient time should affect perceptions of local radio stations and their relative bias, given that the debates put substantial emphasis on providing equal time to all candidates. If these effects on perceived bias also affect perceptions towards the media more broadly, then M6 follows.

Fourth, we consider whether the intervention changed the perception of citizens about elections themselves, given it might have left to a shift toward more programmatic issue and away from vote-buying strategies. We hypothesize that individuals in districts where debate broadcasting was intensive in the weeks before the election have more positive attitudes about the contribution of elections to select

better candidates. Regressions take the form of equation 3.

M7. Individuals believe that the elections allow citizens choosing candidates that are more competent and better represent their interest in intensively-rebroadcast districts: $\beta_3 > 0$ in equation 3.

Outcome variables come from section 5.9.

7.6 Heterogeneous effects and more powerful instruments

We anticipate several sources of heterogeneity. We have outlined some of these – centered on debate performance – but other pre-treatment variables seem natural candidates. First, since we expect the relative radio coverage of respondents (as proxied by their polling places) to be of relevance, we anticipate that the interaction of all our treatments with radio coverage is likely to exhibit a stronger effect.

H1. All effects are stronger when the district coverage of the rebroadcasting station is greater.

Radio coverage is determined using geolocations of radio towers, data on their transmitting power, and radio propagation software. Secondly, we anticipate within-district variation based on when different respondents listen to the radio: effects should be stronger, within a district, for those who listen to the radio at the hours when the rebroadcasts took place.

H2. All effects are stronger when respondents listen to the radio at hours where more rebroadcasts happened.

Thirdly, we expect the invitation intervention to differentially affect the participation decision of the incumbent and other top candidates for reasons outlined in sections 2 and 3: they face higher campaigning opportunity costs from debate participation and larger potential downside from performing poorly in the debate. This leads to H3:

H3. All effects are stronger for incumbents and top candidates.

If the treatment has differential effects in this way, then more incumbents and top candidates should be expected to show up under treatment, which increases exposure to their policy platforms. As such, we should expect all outcomes relating, at least, to knowledge about candidates, competence, and voting behavior to be differentially strong.

Lastly, while above we mention that we intend to use such variation to possibly instrument for the quality of the candidates that show up at the debate, if this not needed we could exploit the assignment of candidates to incumbent debate, the first debate in the district, and the number of candidates in the corresponding debate while controlling flexibly for the number of candidates in the race, to construct a more statistically powerful instrument than assignment to the intensive-invitation treatment.

7.7 Extrapolating effects

Finally, we intend to explore approaches to extrapolating the treatment effects we estimate. The debates initiative attempted to reform several political mechanisms at once: both the supply of credible political

information by candidates, and the broadcasting of political information. Given this reform-based character, we intend to extrapolate our reduced form estimates and estimate what would have happened if the initiative had lacked the intensive interventions that were randomly added to estimate their treatment effect, the aggregate effect of the intensive interventions, and simple counterfactuals of what might have happened if the additional intensive treatments had been used in all districts. To estimate these cases we intend to just use simple linear (or higher order polynomial) extrapolations to provide back-of-the-envelope estimates. These estimates are of great policy relevant, and the ultimate interest of the evaluation of this intervention.

References

- Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of The Nazis in Prewar Germany." *Quarterly Journal of Economics* 130(4):1885–1939.
- Afrobarometer. 2015. "Afrobarometer Data, Liberia, Round 6, 2014." .
URL: <http://afrobarometer.org/data/320>
- Arias, Eric. 2016. "How Does Media Influence Social Norms? A Field Experiment on the Role of Common Knowledge." Working paper.
- Arias, Eric, Pablo Balán, Horacio A. Larreguy, John Marshall and Pablo Querubín. 2017. "Malfeasance Revelations, Coordination Through Social Networks and Electoral Sanctioning: Experimental Evidence from Mexico." Working paper.
- Bidwell, Kelly, Katherine Casey and Rachel Glennerster. 2016. "Debates: Voting and Expenditure Responses to Political Communication." Working Paper.
- Bowles, Jeremy, Horacio Larreguy and Shelley Liu. 2017. "How Weakly Institutionalized Parties Monitor Brokers in Developing Democracies: Evidence from Post-conflict Liberia." *Working Paper* .
- Brollo, Fernanda, Tommaso Nannicini, Roberto Perotti and Guido Tabellini. 2013. "The political resource curse." *The American Economic Review* 103(5):1759–1796.
- Ferraz, Claudio and Frederico Finan. 2008. "Exposing corrupt politicians: The effects of Brazil's publicly released audits on electoral outcomes." *Quarterly Journal of Economics* 123(2):703–745.
- Internews Liberia. 2016. "National Radio Mapping in Liberia." .
- Izama, Melina Platas and Pia Raffler. 2016. "Meet the Candidates: Information and Accountability in Primary and General Elections." .
- Larreguy, Horacio A., John Marshall and Jr. Snyder, James M. 2017. "Publicizing malfeasance: When the Media Structure Facilitates Electoral Accountability in Mexico." Working Paper.
- Liberia Accountability and Voice Initiative. 2016. "NRM Information Gathering Exercise." .
- Olukotun, Ayo and Sharon Adetutu Omotoso. 2017. *Political Communication in Africa*. Springer.
- USAID. 2015. "Liberia Electoral Access and Participation (LEAP) Survey." *USAID* .
URL: http://pdf.usaid.gov/pdf_docs/PA00KJW1.pdf
- Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." *Quarterly Journal of Economics* 129(4):1947–1994.

	β_1	p-value	β_2	p-value
Planned week of debate	0.134	0.472	-0.248	0.175
Number of candidates (2017)	-0.030	0.967	0.879	0.213
Incumbent running (2017)	-0.038	0.582	-0.075	0.265
Number of polling places (2017)	0.106	0.953	0.743	0.678
Registered voters (2017)	312.627	0.837	1563.354	0.297
Log registered voters (2017)	0.019	0.727	0.052	0.330
Projected number of debates	0.031	0.770	0.195	0.071
Vote share of 1st place (2011)	-0.009	0.702	-0.007	0.733
Vote share of 2nd place (2011)	-0.009	0.309	-0.011	0.226
Vote share of 3rd place (2011)	-0.007	0.259	-0.004	0.559
Vote share HHI (2011)	-0.014	0.470	-0.011	0.513
Number of polling places (2011)	0.098	0.955	-0.193	0.911
Turnout (2011)	-0.004	0.591	0.004	0.549
Registered voters (2011)	-1231.767	0.573	-1648.931	0.447
Share candidates who ran in 2011	0.020	0.340	0.017	0.421
Population (2008)	1132.649	0.711	2431.938	0.425
Share urban (2008)	-0.047	0.501	-0.087	0.208
Share male (2008)	0.001	0.629	-0.002	0.561
Average age (2008)	-0.041	0.738	0.114	0.346
Share Christian (2008)	0.029	0.336	-0.044	0.170
Share resettled (2008)	0.050	0.008	0.003	0.883
Share literate (2008)	-0.021	0.277	-0.012	0.525
Share employed (2008)	0.016	0.340	0.006	0.735
Log population (2008)	0.016	0.864	0.069	0.478
Population density (2008)	-0.000	0.517	-0.000	0.770
Log pop density (2008)	-0.186	0.557	-0.201	0.523
Registered voter density (2017)	-0.000	0.561	-0.000	0.926
Log registered voter density (2017)	-0.184	0.648	-0.218	0.588
Number of radio stations (2016)	-0.050	0.839	-0.214	0.385
Number of cell towers (2016)	-0.322	0.620	0.114	0.860
Number of health facilities (2017)	-0.015	0.983	1.713	0.013
Number of police stations (2017)	0.193	0.595	0.411	0.271
Length of roads (2016)	29.521	0.223	28.239	0.236
Roads per sq km (2016)	-0.067	0.866	0.253	0.529
Conflict events (2017)	0.139	0.123	0.012	0.889
GSM coverage (2015)	0.003	0.950	-0.049	0.341
Share owns radio (2015)	-0.022	0.316	-0.014	0.545
Share owns phone (2015)	0.009	0.773	-0.065	0.058
Share gets news from radio often (2015)	0.008	0.702	-0.026	0.252
Share 0-18 (2008)	-0.001	0.700	0.001	0.781
Share 18-30 (2008)	0.003	0.401	-0.001	0.719
Share 30-50 (2008)	0.001	0.725	-0.003	0.106
Share 50+ (2008)	-0.002	0.402	0.003	0.160
Share primary education (2008)	0.004	0.728	0.009	0.393
Share secondary education (2008)	-0.002	0.785	-0.008	0.372
Share tertiary education (2008)	-0.001	0.551	-0.001	0.548
Radio coverage (2016)	0.243	0.825	-1.429	0.214
Radio partner staff	-1.696	0.061	0.710	0.456
Radio partner news shows	-0.019	0.947	0.203	0.509

Table 1: Balance on pre-treatment covariates (Equation 1)

	β_1	p-value	β_2	p-value	β_3	p-value
Planned week of debate	-0.105	0.708	-0.474	0.049	0.140	0.579
Number of candidates (2017)	1.659	0.081	2.474	0.007	-1.044	0.286
Incumbent running (2017)	0.035	0.720	-0.000	1.000	0.116	0.221
Number of polling places (2017)	1.445	0.616	2.000	0.345	-1.017	0.642
Registered voters (2017)	158.448	0.940	1345.316	0.569	-2008.520	0.381
Log registered voters (2017)	0.011	0.887	0.041	0.611	-0.076	0.356
Projected number of debates	0.275	0.032	0.421	0.002	-0.263	0.066
Vote share of 1st place (2011)	0.026	0.447	0.027	0.413	0.014	0.615
Vote share of 2nd place (2011)	-0.005	0.656	-0.007	0.517	0.021	0.087
Vote share of 3rd place (2011)	-0.001	0.898	0.002	0.802	0.011	0.225
Vote share HHI (2011)	0.028	0.304	0.030	0.273	0.023	0.258
Number of polling places (2011)	2.918	0.275	2.526	0.269	-0.115	0.957
Turnout (2011)	0.004	0.672	0.012	0.245	-0.001	0.926
Registered voters (2011)	5846.860	0.095	5274.000	0.052	2557.555	0.379
Share candidates who ran in 2011	-0.034	0.318	-0.036	0.166	-0.035	0.234
Population (2008)	-1140.662	0.820	110.735	0.977	-3663.986	0.340
Share urban (2008)	0.075	0.451	0.036	0.722	0.135	0.207
Share male (2008)	-0.003	0.479	-0.006	0.136	0.001	0.901
Average age (2008)	0.004	0.981	0.154	0.380	-0.082	0.623
Share Christian (2008)	0.099	0.031	0.024	0.533	0.010	0.844
Share resettled (2008)	-0.016	0.565	-0.063	0.025	-0.052	0.042
Share literate (2008)	0.005	0.861	0.014	0.612	0.034	0.243
Share employed (2008)	-0.023	0.332	-0.033	0.175	-0.020	0.432
Log population (2008)	-0.024	0.883	0.027	0.829	-0.089	0.498
Population density (2008)	0.000	0.963	0.000	0.732	0.000	0.569
Log pop density (2008)	0.381	0.372	0.360	0.403	0.374	0.458
Registered voter density (2017)	0.000	0.989	0.000	0.680	0.000	0.680
Log registered voter density (2017)	0.416	0.471	0.374	0.508	0.387	0.544
Number of radio stations (2016)	0.162	0.644	0.000	1.000	0.268	0.405
Number of cell towers (2016)	0.098	0.898	0.526	0.562	0.191	0.823
Number of health facilities (2017)	-1.455	0.139	0.263	0.780	-1.718	0.042
Number of police stations (2017)	-0.757	0.183	-0.526	0.280	-0.577	0.318
Length of roads (2016)	-51.017	0.187	-51.454	0.133	-56.007	0.121
Roads per sq km (2016)	-0.364	0.507	-0.039	0.940	-0.177	0.765
Conflict events (2017)	-0.141	0.244	-0.263	0.031	-0.141	0.265
GSM coverage (2015)	0.086	0.299	0.032	0.680	0.043	0.586
Share owns radio (2015)	0.015	0.610	0.023	0.383	0.036	0.327
Share owns phone (2015)	0.074	0.113	-0.000	0.994	0.055	0.288
Share gets news from radio often (2015)	0.014	0.687	-0.019	0.521	0.019	0.562
Share 0-18 (2008)	-0.000	0.969	0.002	0.672	0.000	0.943
Share 18-30 (2008)	-0.004	0.296	-0.008	0.089	-0.001	0.803
Share 30-50 (2008)	0.003	0.243	-0.001	0.775	0.002	0.375
Share 50+ (2008)	0.002	0.580	0.006	0.050	-0.001	0.553
Share primary education (2008)	-0.005	0.702	-0.000	0.995	-0.013	0.341
Share secondary education (2008)	0.003	0.767	-0.002	0.878	0.011	0.352
Share tertiary education (2008)	0.002	0.507	0.002	0.546	0.003	0.361
Radio coverage (2016)	2.508	0.160	0.797	0.567	1.103	0.478
Radio partner staff	-0.655	0.445	1.754	0.249	0.980	0.318
Radio partner news shows	0.037	0.927	0.273	0.523	-0.209	0.567

Table 2: Balance on pre-treatment covariates (Equation 2)