

DEBATES

The Impact of Voter Knowledge Initiatives in Sierra Leone

Kelly Bidwell

J-PAL

Katherine Casey*

Stanford GSB

Rachel Glennerster

J-PAL

June 19, 2015

Abstract

Debates between candidates for public office have a rich historical tradition and remain an integral part of contemporary campaign strategy. There is, however, no definitive evidence of whether debates affect actual voting behavior. Limited media penetration implies that the effects of publicizing debates could be more pronounced, persistent and directly linked to electoral outcomes in the developing world. We experimentally manipulate citizen exposure to debates between Parliamentary candidates in Sierra Leone to measure their impacts on, and the interconnections between, voter behavior, campaign spending, and the performance of elected politicians. We find evidence of strong positive impacts on citizen political knowledge, policy alignment and votes cast on Election Day. We then document an endogenous response by participating candidates, who increased their campaign expenditure in communities where videotapes of the debates were screened in large public gatherings. A complementary series of individual treatment arms unpacks the different types of information delivered by the debates, and finds evidence that voters respond to both candidate charisma and "hard facts" about policy stance and professional qualifications. Lastly, we find longer term accountability effects on elected MPs, where participation in debates led to higher levels of constituency engagement and development expenditure during their first year in office.

*Corresponding author: kecasey@stanford.edu. This draft remains preliminary and incomplete. This research would not have been possible without the collaboration of Ambrose James and Search for Common Ground; and Innovations for Poverty Action and their Freetown team. We are grateful to the National Electoral Commission of Sierra Leone and members of our expert panel for their collaboration. We are grateful for comments from Maggie Penn, Cyrus Samii and Francesco Trebbi. We thank Allyson Barnett, Fatu Emilia Conteh, Nick Eubank, Abdulai Kandeh, Agnes Lahai, Osman Nabay, Isaac Nwokocha, Katie Parry and Catherine Wright for excellent research assistance. Seminar participants at Bristol University, the Center for Effective Global Action, Experiments in Governance and Politics, International Growth Centre (IGC) Growth Week, Political Institutions and Economic Policy Conference at Princeton, Stanford GSB, University of British Columbia, University of California Berkeley, University of Michigan, Wellesley and the Working Group on African Political Economy provided insightful comments. We gratefully acknowledge financial support from the Governance Initiative at J-PAL, IGC, National Bureau of Economic Research and Stanford Institute for Innovation in Developing Economies. All errors are our own.

1 Introduction

Debates between candidates for public office have a rich historical tradition and remain an integral part of campaign strategy in contemporary politics. In the United States, influential candidate debates include the Lincoln-Douglas Senatorial debates of 1858 and the first televised Presidential debate between Kennedy and Nixon in 1960. More recently, the first Obama-Romney debate of 2012 generated widespread criticism of the President’s performance and much speculation about its implications for his re-election prospects (e.g. Shear 2012). Such debates constitute significant campaign events: a large share of the voting public tunes in to view them; they generate a flurry of media commentary and analysis of candidate performance; and pundits pore over polling data to assess their effects on public opinion. There is, however, no definitive evidence of whether debates have any real impact on subsequent vote shares. Moving to the developing world, the markedly lower penetration of mass media suggests that the effects of publicizing debates could be more pronounced, persistent and more directly linked to electoral outcomes. Given the paucity of reliable political information in these markets, allowing candidates to stand on equal footing and express their views on key policy issues could further facilitate the election of more competent candidates. And, by creating a public record of their pre-election commitments, candidate debates could enhance the subsequent accountability of elected officials. To evaluate these claims, we experimentally manipulated citizen exposure to debates to measure their impacts on, and the interconnections between, voter behavior, campaign spending, and the performance of elected politicians in the context of the 2012 Parliamentary elections in Sierra Leone.

We worked with the nonpartisan, civil society organization Search for Common Ground to host, film and disseminate structured, inter-party debates in fourteen competitive constituencies in the lead up to Election Day. The paper makes four contributions. First, we randomly allocated a “road show” across polling centers that screened videotapes of the debates in large public gatherings to capture effects on voters. We document strong positive impacts of watching debates on political knowledge, policy alignment and importantly, actual votes cast. Second, we document an endogenous spending response by participating candidates who increased their campaign expenditure in communities where these screenings were held. Documenting this indirect result makes a small advance from the typical partial equilibrium orientation of field experiments towards capturing a more general equilibrium effect. Third, individual-level treatments unpack the causal mechanism driving the observed changes in voter behavior, isolating the role of more superficial attributes such as candidate charisma from that of “hard facts” about policy stance and professional qualifications. Fourth, we tracked the performance of winning Members of Parliament (MPs) over their

first year in office to evaluate whether the publicity of the debate screenings eases candidate commitment problems and enhances accountability. We document large positive impacts on constituency engagement and development expenditure by elected officials. Two additional aspects of the research are worth noting: the experiments were in part designed to test theoretical predictions from our earlier work; and the empirical analysis is governed by publicly registered pre-analysis plans.

Large groups of voters were exposed to films of the MP candidate debates via a mobile cinema that visited 112 of 224 randomly selected polling centers in the five weeks before the election. We find strong positive impacts of watching the debates on voter knowledge—of politics in general, and of specific candidate attributes and policy stances; improved alignment between voter policy preferences and those of their selected candidate; and greater voter openness to candidates from all parties. Importantly, these gains in voter knowledge translate into actual changes in votes cast, where we document a five percentage point average increase in vote shares for the candidates who performed best during the debates. This effect is positive and significant in both our exit poll data and in the National Electoral Commission’s official voting returns. In the context of strong historical ties between ethnic groups and political parties, we find no net impact of debates on voting across ethnic-party lines, however document heterogeneous effects along the dimension of ethnic loyalty that are consistent with theoretical predictions from Casey (forthcoming).

As one might expect of any intervention that alters voter behavior on a nontrivial scale, we next document an endogenous campaign response by candidates who participated in the debates. While the candidates were not informed of which polling centers were assigned to treatment, the debate screenings were large public gatherings in rural areas with limited entertainment options, making their locations fairly easy to track down. We find evidence that candidates increased their campaign effort, as measured by gift giving, the monetary value of gifts, and the number of in person visits, in communities where the screenings were held. The increase in expenditure is consistent with a “swing” voter investment model if the debates made exposed areas appear more competitive, by making expected vote margins narrower and/or more uncertain.¹ Tracking the endogenous response of MP candidates to the road show treatment aims to move one small step from a partial towards a more general equilibrium approach. While not similar in design, the goal of documenting the indirect effects of the experiment in its market is similar in spirit to Crépon et al (2013).

A series of treatment arms administered at the individual level disentangles the different types of information revealed by debates—ranging from charisma to policy stance—to see

¹See Lindbeck and Weibull 1987, Dixit and Londregan 1996, 1998, and Bardhan and Mookherjee 2010; and Casey forthcoming for application to ethnic politics.

what drives the observed changes in voter behavior. Specifically, some voters were exposed to brief “get to know you” videos of the candidates speaking informally about themselves and their hobbies, which capture persona but not professional or policy information. Others listened to a radio report or journalistic summary of the debate, which articulated all the “hard facts” about policy stances and professional qualifications that arose during the debates, but delivered no information about persona. Still others watched the full debate on a tablet device. Control respondents were surveyed at the time of treatment and the election, and another group was surveyed only at the election. We find evidence that voters update their views of candidates and policy in response to information regarding persona and hard facts, but that only the combination of the two delivered by the debates moves them into better policy alignment and triggers changes in actual voting choices. Comparing the two sets of controls allows us to further isolate the role of priming (Zwane et al. 2011), where we find that the experience of being surveyed accounts for roughly one third of the overall effect on general political knowledge. Reassuringly, survey priming does not contribute to knowledge of specific candidate attributes or policy stances, nor does it move voters to update their policy preferences or vote choice.

Over the longer term, the publicity of the debate screenings could help solve candidate commitment problems and enhance the accountability of elected officials. By creating a public and enduring record of candidate commitments, debates could make reneging on campaign promises more costly, thereby facilitating greater consistency between pre- and post-election behavior. By informing voters of these commitments and of the resources available to those elected, debates could foster accountability pressure that enhances the performance of winning MPs. To evaluate these interrelated claims, the 14 constituencies that participated in the debate were randomly selected from what we anticipated would be the 28 most competitive races. We then tracked the performance of all 28 elected MPs over their first year in office. We find positive effects on constituency engagement, where for example treated MPs held twice as many public meetings with their constituents, and on the allocation of the constituency facilitation fund, where the value of development expenditures that could be verified in the field was 2.5 times greater for treated than control MPs. We find no evidence for effects on attendance or participation in Parliamentary sittings, nor on consistency in promoting the MP’s priority sector.

There is a large literature concerning the impact of debates in American politics (see for example, Jamieson and Birdsell 1990 or Hellweg, Pfau and Brydon 1992 for review). Much of this work is limited to panel opinion polls, with the familiar identification challenges (Prior 2012). The experimental evidence is inconclusive: one study finds that televised debates impact voter assessment of candidates (Fridkin et al. 2007), while two others find no mean-

ingful effects on political attitudes (Wald and Lupfer 1978) nor opinions (Mullainathan et al. 2010). Additional lab experiments find mixed evidence about how the medium of debate delivery—via television versus radio—affects voter evaluation of candidates (McKinnon et al. 1993, Druckman 2003). Our individual-level treatments delivered via tablet device contribute to this line of research by testing for the impacts of debates in an information poor political environment; unpacking voter responses to multiple different slices of information delivered by debates; and documenting effects of debate exposure on actual votes cast. We view the scale and intensity of the polling center level public screenings as a new contribution. Interestingly, we find an effect size of these group screenings on vote shares that is markedly similar to what Gerber et al. (2011) document in a Texan gubernatorial race for the most intense “dose” of televised campaign advertising on voting intentions. The effects in the U.S., however, dissipate very rapidly, reduced to zero in a matter of days, whereas the impact of our debates persisted over a range of one to five weeks after treatment exposure and affected choices on Election Day. In Italy, Kendall et al. (2015) evaluate campaign messages about incumbent valence delivered by telephone in the week immediately preceding a mayoral election, and find similarly sized effects on vote shares. We are not aware of any other study that links debates to endogenous campaign responses by candidates, nor to accountability effects on elected officials.

In the field of development economics, our approach of working with political candidates in the course of their actual campaigns follows in the tradition of Wantchekon and co-authors.² They find that public deliberation between a single party’s representative and constituents decreases the prevalence of clientelism and increases electoral support for the participating party in Benin and the Philippines. We instead focus on the interaction between rival candidates from different parties, where the head-to-head debates were designed to reveal information about the relative quality and policy differences between candidates. Testing the efficacy of debates further contributes to the literature exploring the impacts of information on voting. Ferraz and Finan (2008) and Banerjee et al (2011), among others, show that providing specific information about incumbent performance and candidate qualifications can have large effects on voting. Debates are distinctive in that they provide more general and comprehensive information about candidates, including information about persuasion and charisma, which can be considered productive attributes of an effective legislator. Moreover, if information about an arguably sufficient competence statistic—like corruption—is not available, the generality of debates could further be important for two reasons. From a theoretical perspective, comprehensiveness eases concerns that increasing

²See Wantchekon (2003), Fujiwara and Wantchekon (2013), and Wantchekon, Lopez-Moctezuma, Fujiwara, Pe Lero and Rubenson (2015).

transparency along one dimension will simply reallocate politician effort towards those more observable actions, regardless of their impact on welfare (e.g. Liessem and Gersbach 2003 on multi-tasking, or Cranes-Wrone et al 2001 and Prat 2005 on pandering). Pragmatically, it makes it harder for politicians to unravel the impact of the intervention: for example, it is easier for them to discredit a scorecard-style information campaign (Humphreys and Weinstein 2012) than a video of their own public statements.

The rest of this paper is structured as follows. Section 2 explains the institutional context, research design and econometric specifications. Section 3 discusses evidence for treatment effects on voters, candidates and elected officials. Section 4 explores survey priming, treatment effect heterogeneity and social mobilization. Section 5 concludes with policy considerations.

2 Context and Research Design

2.1 Institutional Setting

Sierra Leone has 112 Parliamentary constituencies, which are single member jurisdictions elected by first-past-the-post plurality. The winning MP represents the local area, containing approximately 40,000 residents, in the national legislature. In these elections, the ethnic composition of the voters in a given constituency predicts the corresponding party vote shares with remarkable accuracy. These correlations arise from historical ties between the All People’s Congress (APC) party and the ethnic groups in the North, most prominently the Temne; and between the Sierra Leone People’s Party (SLPP) and tribes in the South, most prominently the Mende (see Kandeh 1992). As an example of the contemporary strength of these ties, in the control group of this study, 89 percent of citizens reported voting for the MP candidate from the party that is historically associated with their ethnic group.

Casey (forthcoming) suggests that one reason for the strength and persistence of these ethnic-party ties is that voters have little alternative information on which to base their vote. If true, then providing citizens with better information about the candidates themselves could broaden their decision calculus, potentially making voting more responsive to characteristics like professional qualification, past performance or persuasiveness. Yet how to deliver such information in a way that is credible, engaging, and accessible to voters who are predominantly poor, illiterate and geographically remote remains an open question. It is out of this challenge that our interest in debates as a vehicle to deliver comprehensive information about candidates initially arose, although the study quickly expanded into a broad exploration of debates and their effects on political markets. In this context, candidate debates remain rare but not unheard-of: Presidential debates were held before the 2007

and 2012 elections, however in no debate did both major party candidates participate. Our dissemination vehicle, via mobile cinema, was certainly novel. This paper explores whether and how debates affect voting behavior, how such changes might in turn affect the allocation of campaign expenditure by candidates, and over the longer run, whether debates can enhance the electoral accountability of elected officials.

2.2 Treatments and Randomization

Before the 2012 Parliamentary candidates were officially announced, we randomly selected 14 MP constituencies from what we estimated would be the 28 most competitive races, stratifying on the degree of ethnic-party bias favoring one party over the other. (All randomizations were done on a computer.) Our NGO partner, SFCG, then invited candidates from the three largest parties—the APC, SLPP, and the latter’s splinter party, the People’s Movement for Democratic Change (PMDC)—who were contesting a given seat to participate in a debate. No other parties won seats in the previous election, and these three parties respectively held 59, 39 and 9 percent of the current seats in Parliament.

Each of the fourteen debates followed a standardized format. The SFCG moderator opened the debates by introducing the candidates and explaining the basic roles and responsibilities of office. A casual “get to know you” section followed, where the candidates spoke informally about where they were from, their family and hobbies. Then five national policy questions were posed, where each candidate was allowed two to three minutes to respond. The first policy question concerned the candidate’s top priority for additional government spending. The second covered plans for how to spend the constituency facilitation fund (CFF), which is an untied 43.8 million Leones (approximately US\$ 11K) grant given annually to each MP for transport to and development of his or her constituency. The third issue asked for the candidate’s strategy to uplift the youth, where “youth” is defined by the government as 18 to 35 year old adults. This demographic segment faces high unemployment and their historic disenfranchisement and frustration were seen by many as a contributing factor to the country’s civil war (1991 to 2002). Fourth was whether the candidate, if elected, would vote in favor of or against the Gender Equity Bill (GEB), a 30% quota for women’s representation in government that was introduced but never voted on by the previous Parliament. The last national policy question asked for the candidate’s assessment of the implementation of free healthcare (FHC), a major initiative by the incumbent government to provide free care to children under five and pregnant or nursing women. Each debate closed with two local policy questions, which varied by constituency and were tailored to prominent issues in the local area. All debates were conducted in Krio, Sierra Leone’s *lingua franca*.

Within the fourteen constituencies selected for participation in the debates, we first allocated polling centers to the group screening treatment and control arms. All citizens had to register anew for this election, and the polling centers—typically a primary school or community center—are where they went to register and then later to vote. This sample drew in 224 polling centers that had fewer total registered voters (471 on average) and were located further away from their nearest neighboring polling center (2.4 miles on average) than the population in general. SFCG took videotapes of the debates on a “road show” to 112 of these polling centers, selected randomly. In the eight constituencies where there were a sufficient number of polling centers left over, we randomly allocated 40 of the remaining larger and closer together centers into the individual-level treatment group. Note that the individual treatment arms were thus administered in a completely separate set of communities from the public screenings. A few months before administering any intervention or survey, we conducted a household listing of registered voters in all 264 polling centers to develop the sampling frame for individual respondents.

The “road show” or mobile cinema treatment at the polling center level consisted of an evening showing of the video of the relevant debate projected at a convenient public place, usually on the side of the polling center itself, in the weeks leading up to the Election. Typical protocols for these screenings were as follows: host polling center and satellite communities were notified in advance and invited to attend the screening; 25 randomly selected residents (using data from the earlier listing exercise) were provided a small incentive (10 cooking spice cubes) to attend the screenings; the video was played once in a pause and play format that inserted translation into the relevant local language after each question; the video was played a second time with or without translation; and a secondary screening was held in the largest accessible satellite community earlier in the day, in most cases without translation. A total of 112 primary and 85 secondary screenings were held, and we estimate that roughly 19,000 people were thereby exposed to one of the debates.

At the time of screening in treated polling centers, some of the 25 respondents who received attendance incentives were also surveyed, where specifically: i) 12 completed surveys both before and after the screening; ii) 4 completed only after screening surveys; and iii) the 9 remaining were not surveyed but were contacted only to deliver the incentive. We later conducted exit polls on Election Day and the days immediately after in all 224 treatment and control polling centers. To avoid any differential attrition or selection across treatment assignment, the 5,600 exit poll respondents were drawn from the original household listing in both treatment and control polling centers and surveyed at their residence. In what follows, we will thus be estimating intention to treat effects, where 82% of exit poll respondents indicated that they had attended a debate screening, as did 4% of those in the control

group.

Within each of the polling centers assigned to individual-level treatments, households were divided into those with only female registered voters, only male, and both male and female registered voters (based on the earlier household listing exercise). We randomly assigned treatment arms to households within each of these bins, and randomly selected respondents within each household to receive the individual-level treatments and/or survey(s). The treatment arms at the individual level were as follows: (i) debate treatment, where individuals were shown the exact same debate screened in polling centers on a personal handheld device; (ii) “getting to know you” treatment, where individuals were shown a short video of the same two candidates speaking informally about their hobbies and interests; (iii) “radio report” treatment, where individuals listened to a journalistic summary of the main policy positions articulated by the candidates during the debates; (iv) surveyed control, where individuals were given the same survey as the one that accompanied treatments i to iii, but were not shown any media; and (v) pure control, where individuals were not surveyed until Election Day, and whose only contact with the research team at time of treatment implementation was to record basic demographics. A sixth treatment arm was administered involving a lab-in-the-field experiment that exposed voters to photos and 20 second video clips of candidates (analyzed in our related work). This treatment did not relay any political information and the media snippets did not cover any candidates from the voter’s own constituency, and is thus grouped with the control arms.

There were 400 individuals assigned per treatment arm and 600 in the surveyed control group. Unlike for the polling center level intervention, the exact same respondents who participated in the individual treatment arms were relocated in the exit polls. As we had perfect compliance and minimal attrition (at 6%, which does not vary by treatment assignment), average treatment effect estimates for the individual treatment arms are comparable to treatment on the treated effects.

2.3 Hypotheses and Econometric Framework

We registered the first and main pre-analysis plan (PAP) governing this analysis with the Abdul Latif Jameel Poverty Action Lab on November 20, 2012 before fieldwork for the exit poll, which is the primary source of data for this analysis, was completed. We later migrated the PAP when the American Economic Association’s randomized control trial registry opened, where our entry can be found here: <https://www.socialscienceregistry.org/trials/26>. To establish an iterative process that used earlier data analysis to inform subsequent analysis, and in light of data collection efforts spanning 18 months, we later lodged additional plans

and revisions with time stamps to the registry (explained further below). Overall, the plans establish five research domains with hypotheses under each domain; group outcomes under these hypotheses; and specify the econometric framework including subgroup analysis, dimensions of heterogeneous effects, and which tests are one-sided and in which direction.³

The PAP commits to reporting treatment effects for all individual outcomes as well as mean effects indices by hypothesis, and then adjusting for multiple inference across hypotheses within a domain and across outcomes within a given hypothesis. Compilation of the mean effects index follows Kling, Liebman and Katz (2007) to first orient each individual outcome so that larger numbers imply better outcomes, translate each into standard deviation units with reference to the mean and standard error of the control group, and then compute the equally weighted average of all transformed outcomes under a given hypothesis.⁴ Following Anderson (2008), we then apply family wise error rate (FWER) adjustments at the hypothesis level, which strongly control the probability of making any type-I error; and apply false discovery rate (FDR) adjustments at the individual outcome level, which control the expected proportion of rejections that are type-I errors. Note that we do not adjust across research domains (e.g. across voters in the polling-center versus individual-level treatment arms), as each domain concerns a distinct sample: covering different actors, datasets and/or randomizations.⁵

The PAP lists the following hypotheses for the first research domain (A), which concerns the effects of the polling center-level debate screenings on voters:

A1. Exposure to debates increases political knowledge and leads to more informed voting, including (i) general political knowledge; (ii) knowledge of individual candidate attributes; and (iii) candidate policy stances

A2. Exposure to debates increases policy alignment

A3. Exposure to debates increases vote shares for the candidate who performed the best in the debate

A4. Exposure to debates increases the willingness to vote across party lines

A5. Exposure to debates enhances voter openness to other parties

Secondary hypotheses: Exposure to debates (i) mobilizes the public and leads to greater turnout; (ii) increases the perceived legitimacy of elections; and (iii)

³See Casey, Glennerster and Miguel (2012) for discussion of PAPs.

⁴Missing values for individual index component outcome measures imputed at the random assignment group mean (see KLK 2007 footnote 11).

⁵Note that the word “domain” is often used to refer to different groups of outcomes tested on the same dataset. Our “domains” are thus quite distinct from that usage and imply a much stronger degree of separation between tests.

increases interest in politics

Analysis of treatment effects for domain A takes the form:

$$Y_{ipc} = \beta_0 + \delta T_{pc} + \mathbf{X}'_{ipc} \boldsymbol{\Pi} + \mathbf{Z}'_{pc} \boldsymbol{\Gamma} + \mathbf{W}'_{ipc} \boldsymbol{\Psi} + c_p + \varepsilon_{ipc} \quad (1)$$

where outcome Y (i.e. vote choice) is measured for individual i registered in polling center p within Parliamentary constituency c ; T is an indicator variable equal to one if the polling center received the debate group screening treatment; \mathbf{X} is a vector of indicator variables that denote the stratification bin from which exit poll respondents were drawn (where the bins were constructed by age and gender); \mathbf{Z} is a vector of indicator variables that denote the stratification bin from which the polling center was drawn (where the bins were constructed by number of registered voters and distance to nearest neighboring center); \mathbf{W} is a set of individual controls determined by a pre-specified algorithm that uses control group data to select the subset of {gender, age, years of schooling, polygamous marital status, farming occupation and radio ownership}⁶ that predicts the mean effects index for a given hypothesis at 95% confidence; c is a set of constituency-specific fixed effects (the level of debate and candidates); and ε is an idiosyncratic error term clustered at the polling center level. The coefficient of interest is δ and recall that we are estimating intention to treat effects. Unless otherwise stated, all tests are one-sided in the direction indicated in the statement of the hypothesis. The PAP further specifies the following primary dimensions of potential heterogeneous effects: (i) competitiveness of constituency; (ii) candidate performance; and (iii) subgroup analysis by gender, age and fluency in Krio; which are discussed in Section 4.2.

For the second research domain (B), the PAP lists only one hypothesis concerning the effects of polling center-level debate screenings on candidates:

B1. Candidate allocation of campaign effort and expenditure is responsive to debate publicity

Here we are interested in whether campaign investment complements or substitutes for treatment allocation, and thus conduct two-sided tests. In terms of theoretical motivation, if debates influence vote shares in a way that makes the races appear more competitive, then they could attract greater resources to areas where the debates were screened. This would be consistent with a “swing voter” model of political redistribution (Lindbeck and Weibull 1987), which Casey (forthcoming) finds evidence for in Sierra Leone. Recall that while we

⁶Interest in politics was removed from the pre-specified set as it is potentially endogenous to treatment.

did not inform the candidates of which polling centers were assigned to treatment or control, the screenings were large public events whose locations would not have been difficult to track down after they occurred. This is thus an endogenous response of candidates to the polling center-level treatment assignment. Treatment effects in domain A concerning voters thus capture the combination of exposure to debate and the campaign response. (By contrast, individual-level treatments administered under domain D below capture a “pure” debates effect absent any campaign response.) The econometric specification is exactly the same as presented in Equation (1), save the outcomes are linked to individual candidates: e.g., an outcome Y (i.e. receiving a gift) is measured for individual i in relation to candidate m where the individual is registered in polling center p within Parliamentary constituency c .

The PAP repeats the hypothesis above, only now applied to political parties more generally as opposed to individual candidates, to establish the third research domain (C). Data for this domain were collected in a community-level survey that accompanied the voter-level exit polls, implying that there are many fewer (by an order of magnitude) observations for this analysis than for domain B. Survey questions here do not distinguish gifts from different party representatives, and instead reference any party official or candidate for office, where the offices include President, MP, Local Councillor, and Local Council Chair, all of which were contested during the single General Election under study. The hypothesis covers additional outcome measures, like political rallies and number of posters displayed, that apply only at the community-level.

For the fourth research domain (D), we registered a separate PAP to govern the analysis of the individual treatment arms. The hypotheses and outcomes are the same as those specified for domain A above, but we are now interested in the absolute treatment effect of each of the three treatment arms (debate, get to know you and radio report) compared to the control group, as well as the net or relative effect of each treatment arm compared to the other treatments. Analysis of individual-level treatment arms takes the form:

$$Y_{ihtpc} = \beta_0 + \delta T_{htpc} + \mathbf{X}'_{hpc} \boldsymbol{\Pi} + \mathbf{Z}'_{pc} \boldsymbol{\Gamma} + \mathbf{W}'_{ihtpc} \boldsymbol{\Psi} + c_p + \varepsilon_{ihtpc} \quad (2)$$

where outcome Y (i.e. vote choice) is measured for individual i living in household h assigned to treatment arm t registered in polling center p located in Parliamentary constituency c ; T is a dummy variable indicating assignment to treatment arm t ; \mathbf{X} is a vector of indicator variables that denote the stratification bin from which the household was drawn (where the bins were determined by the gender composition of registered voters); and \mathbf{Z} , \mathbf{W} , c and ε remain as defined in (1). For each treatment arm, the coefficient of interest is δ_t , the average treatment effect for treatment t compared to the control group. The control group is defined

as respondents in both the surveyed and “pure” control arms as well as participants in the sixth lab-in-the-field arm (who received no information about these MP races). We further test a series of hypotheses about the relative effects of the different treatment arms that take the form $\delta_t \neq \delta_{-t}$. Tests of average treatment effects are one-sided in the direction of the hypothesis statement, and tests of relative effects are two-sided.

The fifth and final research domain (E), explores medium term accountability effects of the debate treatment on the candidates who won the seat. This analysis operates at the highest level of aggregation, where we randomly allocated 14 of the 28 constituencies into debate participation. We surveyed all candidates in the 28 constituencies pre-election, and tracked the 28 winning MPs over their first 18 months in office. Their performance outcomes were drawn from Parliamentary administrative records, MP self-reports, and extensive fieldwork in their home constituencies. There are four hypotheses in this domain:

- E1. Accountability pressure of constituent exposure to debates is expected to increase the activity and engagement level of elected MPs
- E2. The publicity of the debates helps solve the candidate commitment problem and makes their post-election behavior in Parliament more consistent with their pre-Election promises
- E3. Accountability pressure of constituent exposure to debates is expected to increase post-election engagement with constituents
- E4. Accountability pressure of constituent exposure to debates is expected to increase development expenditure under the CFF

The econometric specification here is:

$$Y_{ic} = \beta_0 + \delta T_c + \mathbf{X}'_i \boldsymbol{\Pi} + \lambda_c + \varepsilon_{ic} \quad (3)$$

where Y_{ic} is outcome for MP candidate i who won the seat for constituency c , T_c is an indicator signaling that the constituency was assigned to the debates participation treatment, X_i is a vector of MP-level controls {gender, public office experience} chosen for their contribution to increasing R^2 in the control group data, and λ_c are fixed effects for the randomization strata used in the constituency-level assignment (three bins of raw ethnic-party bias). Tests are one sided in the direction of better performance. Given the small sample, we do not have power to adjust for multiple inference in this domain. The PAP identifies Hypothesis E4 concerning CFF spending as the one hypothesis in this domain where we have the strongest *a priori* interest and thus where the per comparison p -value is most relevant.

As referenced above, we in total lodged three PAPs and two updates in an iterative process that tracked the sequential analysis of the many datasets we collected. The important thing to note is that the hypotheses and outcome measures for domains A, B, C and D were all established with the first plan in November 2012 before collection of the primary data, the exit poll, was completed; and those for domain E were lodged in June 2014 before the constituency-level fieldwork tracking the activity of winning MPs was completed. Building on these, one additional plan sets out the mechanisms related to the individual treatment arms analysis; one update transparently records revisions to the first polling center-level PAP as analysis of earlier data collection efforts refined plans for subsequent analysis; and one update refines the specific indicators for elected MPs in domain E after analysis of the control group data. All revisions are documented in the online appendix and AEA registry. There were three substantive revisions to the first PAP worth noting here. First, we “demoted” the hypothesis about effects on turnout from primary to secondary after official election results were published revealing very high (87.3%) turnout rates, implying that we would have limited power to detect treatment effects. Second, we combined two hypotheses in the initial plan—policy alignment and policy persuasion—into one single hypothesis, as they capture different mechanisms leading to the same observable outcomes. Third, we added analysis of survey priming, however its earlier omission was a simple oversight as the original research design explicitly includes surveyed and pure controls to capture these effects. Throughout the rest of the paper, we clearly indicate a small number of analyses that were not pre-specified and should be thus considered exploratory rather than confirmatory in nature.

3 Estimated Treatment Effects by Research Domain

3.1 Effects of Debate Group Screenings on Voters (Domain A)

Table 1 presents summary results at the hypothesis level regarding effects of group exposure to the debates via the mobile cinema road show. We find evidence for strong positive treatment effects for four out of the five primary hypotheses concerning impacts on voter behavior. For the first hypothesis, the treatment coefficient for the mean effects index in row 1 suggests that watching debates increases political knowledge on average by 0.281 standard deviation units (standard error 0.028) across the 20 individual outcome measures included. The second column displays the per comparison p -values based on one-sided tests in the direction pre-specified by the hypothesis statement in the PAP. This per comparison, or “naïve,” p -value is appropriate for any researcher with an *a priori* interest in the specific hypothesis presented (see discussion in Kling, Liebman and Katz 2007). For political knowl-

edge, the p -value indicates significance at greater than 99% confidence. The final column adjusts p -values to account for the fact that we are simultaneously testing five distinct hypotheses on the same dataset. These family wise error rate (FWER) adjustments are quite conservative, as may be appropriate for assessing the overall effectiveness of the program and making policy decisions about whether additional funds should be allocated towards scaling up implementation (see discussion in Anderson 2008). This adjustment does not affect the significance of the effect on political knowledge (FWER p -value of 0.000).

To understand what drives the observed increase in political knowledge, Panel A1 of Table 2 unpacks the mean effects index into its individual components, showing results in units natural to the specific outcome. We break this hypothesis into three knowledge subsets covering general political information, candidate attributes and candidate policy stances. The first four outcomes, preceded by (i), capture effects on general political knowledge. In the first row, the first column shows that the proportion of voters in control polling centers who could correctly state the amount in the constituency facilitation fund (CFF) was only 0.034 or 3.4%. Note that answers were coded to correct for a generous range around the actual 43.8 M Leone figure (roughly US\$ 11 K). Column 2 presents a treatment effect estimate of 0.140 (s.e. 0.018) or 14 percentage points, indicating that the proportion of voters who knew the amount in the CFF increased fivefold with exposure to treatment. This effect is highly significant, with a per comparison or “naïve” p -value of 0.000 in column 4. Similarly, in row 2 the proportion of voters who knew who was eligible to receive free healthcare increased by 5.6 percentage points (s.e. 3.3) on a base of 70.6%. Row 3 shows that the treatment effect estimate for correctly stating that the Gender Equity Bill would reserve between 2 and 4 out of 10 seats for women is positive, but not statistically significant. Lastly, in row 4, the number of correctly reported roles and responsibilities of an MP also increased significantly. Note that the statistical strength of these results is largely unchanged when we adjust p -values to control the false discovery rate (FDR) across all 43 primary plus 5 secondary outcomes within domain A (column 5).⁷ While the FDR adjustments generally inflate p -values throughout the table, for knowledge of the GEB quota and some other larger p -values they actually adjust downwards, which can be expected when there are many true rejections in the test set.

The next set of individual outcomes under hypothesis A1 concerns voter knowledge of specific candidate attributes, demarcated by (ii) in Table 2. Here we again find strong positive treatment effects, which are significant at greater than 95% confidence for 6 of the 7 outcomes measured. As some examples, the proportion of voters who could infer which

⁷This adjustment across all outcomes within domain A is actually more conservative than what we specified in the PAP, which was to adjust only across outcomes within each of the five hypotheses.

candidate was better educated rose from 24.3% to 40.2%, and the proportion who knew which candidate (if any) had been an MP in the past increased from 49.0% to 60.1%, both significant at 99% confidence (s.e. 4.4 and 3.2, respectively). Voter knowledge of public office experience and ability to name candidates for all three parties also increased significantly.

For the third and final area of political knowledge, we find evidence that voter knowledge of candidate policy positions also increased markedly. For each of (up to) three participating candidates, on each of three national policy issues, voter ability to correctly place the candidate on the specific policy spectrum increased significantly (at 99% confidence) for 8 of 9 estimates. As some examples, the proportion of voters who could correctly identify the SLPP candidate’s first priority for government spending doubled, from 14.2 to 29.1% (s.e. 2.8); the proportion who knew the APC candidate’s view on whether free healthcare was being well implemented or needed to be significantly reformed rose from 25.2 to 44.9% (s.e. 3.5); and the proportion who knew whether the PMDC candidate would vote in favor of the gender equity bill (GEB) rose from 24.4 to 45.3% (s.e. 5.3).

Together, these results suggest that exposure to debates led to substantial improvements in voter knowledge. Recall that respondents experienced a one to five week lag between exposure to debates and the exit polls, indicating that these gains in knowledge were relatively persistent. The next natural question is thus whether these knowledge gains translated into changes in voting choices on Election Day.

Returning to Table 1, estimates in the second row suggest that voters acted on the gains in policy knowledge to move into better policy alignment with their preferred candidate. Alignment is measured as a match between the voter’s reported policy position in the exit poll and that of the candidate they voted for as expressed by the candidate during the debate. The treatment effect coefficient suggests that debate exposure increased policy alignment by 0.106 standard deviation units (s.e. 0.035) on average across the three major national policy issues discussed during the debates. This effect is significant at 99% confidence for both per comparison and FWER controlled p -values. To provide a sense of magnitude, consider estimates in Panel A2 of Table 2. The empirical match between the voter’s first priority issue and the view articulated by their preferred candidate during the debate increased by 9.0 percentage points (s.e. 3.1) on a base of 42.5%. We find similar effects on alignment with respect to free healthcare, where alignment increased by 9.2 percentage points (s.e. 3.5) on a base of 39.4%. We find no effect for the gender equity bill, although note that alignment was already markedly higher in the control group (61.3%) and there was little divergence in views expressed by candidates during the debates, as only one candidate expressed strong objection to the bill.

What drives this improvement in policy alignment: choosing aligned candidates or per-

suasion and opinion updating? Using first priority issue as an example, alignment improves if: i) voters who prefer education select a candidate who also supports education; and/or ii) voters update their view that education is the most important sector after observing their preferred candidate advocate for education. The former is what one would expect from canonical minimum policy distance voting models (e.g. Downs 1957). By contrast, Abramowitz (1978) suggests that the latter was at work in the Carter-Ford Presidential debates of 1976, where voters adopted their preferred candidate’s view on unemployment policy after watching the two candidates debate the issue. Lenz (2009) further argues that these effects are concentrated among voters who learned the candidates’ positions from the debate. We have two ways to distinguish these channels in our data.

First, we can compare changes in alignment for voters who selected the same party in the previous Parliamentary elections (in 2007), suggestive of persuasion, to those for voters who changed their party choice for this election, suggestive of selecting an aligned candidate. For all three issues, the coefficient on the interaction between voting for the same party in both elections and treatment is not statistically distinguishable from zero. Estimating effects on alignment for the two types of voters separately, the coefficient on treatment is larger in magnitude for the persuasion subsample for first priority issue but smaller for views on free healthcare (all four estimates remain positive and three statistically significant in the subsamples). This suggests that improvements in alignment are roughly comparable for both types of voter, or that both channels are in fact at work. Note, however, that the population size of the group who supported the same party over time is much larger (encompassing 83% of the sample), suggesting that persuasion is the more empirically substantive channel.

As a second strategy to disentangle persuasion from choosing aligned candidates, we can focus on voters from ethnic groups not historically affiliated with either party. As these voters have weaker party attachments, they are presumably less constrained in selecting a candidate based on policy positions. This is not the case empirically: the treatment effect on policy alignment in healthcare for the unaffiliated subsample is not statistically distinguishable from zero (-0.030, s.e. 0.065). By contrast, the same treatment effect for voters who selected their historically aligned party is positive 0.116 (s.e. 0.040) and highly significant. These estimates provide further evidence that persuasion or opinion updating is the empirically dominant channel.

The treatment effect of ultimate interest is on the third hypothesis, where we find significant positive impacts on votes cast for the candidate who performed best during the debates. Specifically, estimates for the mean effect index in row 3 of Table 1 suggest an increase of 0.086 standard deviation units (s.e. 0.043), significant at 97% confidence on a per comparison basis and 92% confidence under FWER adjustment. This index compiles two

measures of debate performance: one determined by the audience and another by our expert panel. Audience judgments were recorded in a survey that immediately followed the implementation of the group-level screening. The expert panel consists of over thirty members of government and civil society who watched the debate videos and scored candidate responses to each debate question. These two sets of evaluations coincided on who performed best in 10 out of the 14 debates. Where they diverged, the expert panel was more likely to pick a less popular candidate, including one from the PMDC, the smallest party that was not very competitive in this election (they won zero seats nationwide).

Table 2 reports treatment effects for these two measures in our exit poll data (primary test) in Panel A3, and in the National Electoral Commission’s (NEC) official polling-center level returns (secondary test) at the bottom of the table.⁸ Note that the correlation between party vote shares measured across the two datasets is .93 for the APC and 0.92 for the SLPP, suggesting that misreporting of vote choice in the exit polls is not a major concern. All four treatment effect estimates for votes for the debate winner are positive, and three are significant at 95% confidence. The estimate that is largest in magnitude is for votes for the candidate that audience members judged to have performed the best, measured in the exit poll data, where we see a 4.9 percentage point (s.e. 2.1) increase in votes for the debate winner. As a benchmark, this effect is comparable to the estimated incumbency advantage of American state legislators (Ansolabehere and Snyder 2002). The corresponding estimate using the official NEC returns is somewhat diluted, to 3.5 percentage points (s.e. 1.7), as expected given that the returns data includes voters from peripheral villages not exposed to treatment. Reassuringly it remains significant at 95% confidence.⁹ Note that vote shares for these candidates were already high, at 71% in the NEC returns for the control group polling centers, indicating that in this set of constituencies, the candidate who was locally popular tended to also perform best during the debates.

For hypothesis A4, we find no evidence for treatment effects on voter willingness to vote across ethnic-party lines. In Table 1, the treatment effect coefficient for the mean effects index is small in magnitude and not statistically distinguishable from zero, as are all three estimates for the associated individual outcome measures in Table 2. How can we reconcile

⁸The NEC sample excludes constituency 15 because the SLPP candidate was disqualified immediately before the Election but his name remained on the ballot, resulting in 48% of ballots cast being deemed invalid (many of which were likely SLPP votes). The winner was eventually determined via the courts. Treatment effect estimates remain largely unchanged when this constituency is included (0.032, s.e. 0.016* for audience best and 0.032, s.e. 0.015* for expert best, N=224).

⁹Our PAP commits to showing estimates when including an additional 29 “pure” control polling centers located in 3 of our constituencies that were randomized out of our study sample. As we defined the randomization strata after their exclusion, we must alter the main specification somewhat to include these extras. Treatment effect estimates remain similar with their inclusion: 2.8 percentage points for both votes for audience and expert best, with one-sided p-values of 0.077 and 0.073, respectively.

a 5 percentage point shift in votes toward the winner of the debate, with no commensurate change in voting across ethnic-party lines? Here exploration of heterogeneous treatment effects along two separate dimensions are informative, although note that neither were included in the PAP so their analysis should be considered exploratory and not confirmatory in nature.

First, the treatment effect on votes for the debate winner are inversely proportional to both the strength of the voters’ ethnic loyalty to that candidate’s party and their population share of the study sample. Specifically, for voters historically aligned with the party of the debate winner, there is no treatment effect (1.6 percentage points, s.e. 1.4) of watching the debate on their vote choice, as presumably they were already planning to vote for that candidate. These voters constitute 81% of the study sample and had baseline rates of 90% voting for the aligned candidate (i.e. debate winner) in the control group. By contrast, roughly 12% of the sample are composed of voters from ethnic groups that do not have strong historical ties to either party. About half (57%) of these voters chose the debate winner in the control sample (as one would expect if they were truly unaffiliated). The point estimate on the treatment effect for this group is substantially larger, at 10.1 percentage points (s.e. 8.4), but not statistically significant (one sided p -value of 0.115). Note that these unaffiliated voters are excluded from the sample for the crossing ethnic-party lines estimate, as they do not have an ethnic choice to move away from. And finally, voters from ethnic groups historically affiliated with the rival party (i.e. the candidate running against the debate winner), represent only 7% of the sample and had a treatment effect estimate of 10.6 percentage points (s.e. 7.5), which is significant at 92% confidence in a one sided test. Thus, while the net effect of voting across party lines is zero, heterogeneous effects on switching one’s vote to the debate winner are concentrated in the small group of voters historically affiliated with the rival tribe, who are indeed moved to cross party lines.

For the second test, we focus on the single constituency where the audience deemed that the “outsider” candidate (who received only 26% of the votes in the control group) was the debate winner. Here the treatment effect on votes for the winner is four times larger than in the full sample (19.1 percentage points, s.e. 11.0) and significant at 94.8% confidence in a one-sided test. These related pieces of evidence are consistent with the theoretical model in Casey (forthcoming), where better information about candidates increases voter willingness to vote across ethnic-party lines, but only in the event that the rival party runs a sufficiently superior quality candidate.

Estimates for the fifth and final hypothesis suggest that exposure to the debates enhanced voter openness to candidates from all participating parties. In Table 1, we see that the treatment effect for the mean effect index is 0.091 standard deviation units (s.e. 0.048),

significant at 97% confidence based on unadjusted p -values and 92% confidence based on FWER adjusted p -values. This index compiles information from 10 point likeability scales, where all five treatment effect estimates in the individual outcomes are positive (in Table 2) and one is statistically significant at conventional levels. While clearly voters updated more positively for some candidates than others, the fact that their opinions rose across the board for all candidates is an important feature for securing candidate participation in future debates.

3.2 Endogenous Response by Candidates and Other Party Officials (Domains B and C)

Domain B explores whether candidates altered their campaign strategy in response to the debates road show, given its strong effects on voters' political knowledge and preferences. Table 3 presents evidence that candidate campaign spending serves as a complement to the publicity of the polling center debates screenings. In the first row, the treatment effect for the mean effects index is 0.103 standard deviation units (s.e. 0.039), which is significant at 99% confidence using a two-sided test. The next nine rows unpack this index into its component measures and present treatment effect estimates for each individual outcome, per comparison or naïve p values, and false discovery rate (FDR) controlled q values that adjust for multiple inference. Three things are worth noting. First, all nine treatment effect estimates – covering candidates from each of three parties and each of three campaign outcomes considered – are positive in sign. These reflect increases in voter reports of having received a gift from the particular candidate, the monetary value of the gift (expressed in logs), and the number of times the candidate was reported to have visited the community, all with reference to the weeks leading up to the Election. Second, the response by candidates from the two major parties, the APC and the SLPP, is roughly proportional when measured as the percentage increase on their base level of spending in control communities. Third, the candidates who appeared to have responded most strongly with complementary spending are those from the PMDC, the third party that generally had less of a chance of winning seats.

What drives this reallocation of campaign effort? One potential explanation is that by equipping voters with greater political knowledge and changing their voting choices, debate screenings made these areas more competitive. This would be consistent with a “swing voter” campaign investment model (see for example Lindbeck and Weibull 1987). Extending the exploratory analysis above (and again noting that this was not pre-specified), the treatment effect on the campaign index is five times larger in the constituency where the “outsider” candidate won the debate (at 0.41 standard deviation units, s.e. 0.16) compared to the other

constituencies in the sample, which is precisely where the debates had the largest impact on the competitiveness of the race. The coefficient on this difference (0.33, s.e. 0.16) is significant at 95% confidence. Note, however, that the coefficient for the remaining constituencies, where the screenings de facto made the races less competitive as the locally popular candidates performed better in the debates, remains positive and statistically significant at 95% confidence (0.08, s.e. 0.04). This can be reconciled with the idea of greater competition if the debates made the vote shares in screening communities more uncertain, as recall that the actual impact of the debates on voting was not revealed until Election Day. This is consistent with the model in Casey (forthcoming), where information increases voter responsiveness to individual candidate attributes, thereby making it harder for candidates to infer vote shares from the ethnic composition of a locality, and thus widening the set of potentially competitive areas.

Appendix Table 1 presents results for domain C, about whether other party officials beyond the candidates directly involved in the debates responded to the publicity of the road show. We find little evidence that other party officials, including centralized party bosses, and candidates for President, Local Councillor and Local Council Chair, altered their campaign strategy in response to dissemination of the MP candidate debates. While the treatment effect for the mean effects index is positive in sign (0.082 standard deviation units), it is not significant at conventional levels (s.e. 0.052 and p -value 0.113). Similarly, while the majority (16 of 21) of treatment effect estimates for the individual outcomes are positive in sign, none are significant at conventional levels. One interpretation of these results is as a pseudo placebo test, where candidates for offices not involved in the debates would not be expected to alter their campaign strategy in response to the road show. This would make sense if the parties did not strongly coordinate campaigns across candidates for different offices, or if the road show was not a salient enough event to justify reallocating campaign support from other party members to support the participating MP candidates. While this seems plausible, we do not want to place too much weight on this interpretation, for two reasons. The sample size for this community-level survey is small (224 communities), so power to reject the null is limited. And, the community survey questions bundled together the campaign efforts of all party officials and candidates for all offices, which includes Parliament, so they do not clearly exclude the MP candidates as one would do for a true placebo.

3.3 Unpacking Causal Mechanisms via the Individual Treatment Arms (Domain D)

Table 4 turns to the series of treatment arms administered to individual voters to explore the relative effects of different types of information conveyed by the debates. Each hypothesis is summarized by its mean effects index, where columns 1 and 2 present the treatment effect estimates and associated per comparison, one-sided p -values for the debate arm, columns 3 and 4 for the “get to know you” video arm, and columns 5 and 6 for the radio report, all with respect to the full set of controls. We show results for political knowledge at both the hypothesis and sub-hypothesis level. Columns 7 through 12 estimate the relative size of treatment effects across the 3 arms, using two sided tests for each two-way combination. In these columns we present both per comparison p -values and FDR q values that adjust across all 24 tests run (three treatment arm combinations by five hypotheses and three sub-hypotheses).

The first row of columns 1 through 6 test for treatment effects on hypothesis A1 concerning political knowledge and suggest that all three arms were effective in transmitting political information. The treatment effect on the mean effect index is positive and significant at 99% confidence for all three arms. Making comparisons across arms, the coefficient for debates and the radio report are more than twice as large in magnitude as that for the get to know you video, differences that are both statistically significant under per comparison p and FDR adjusted q values in Columns 7 and 11. Specifically, the treatment effect for debates is 0.109 standard deviation units (s.e. 0.021) compared to 0.041 (s.e. 0.016) for get to know you, yielding a difference of 0.068 standard deviation units (s.e. 0.022) which is significant at greater than 98% confidence under both a single two-sided test and the FDR adjustment for the fact that it is one of 24 comparative tests run. While the coefficient on debates is slightly larger than that for the radio report, the difference is not statistically distinguishable from zero.

The next three rows separate the knowledge questions into the three sub-hypothesis indices specified in the PAP. The pattern of treatment effects for general political knowledge mirrors that of the hypothesis overall, where all three arms yielded strong positive treatment effects, and the debates and the radio report estimates are larger in magnitude than those for get to know you video. Estimates in row three cover the 7 outcomes concerning knowledge of candidate characteristics, where again, all three estimates are positive and significant, although now the magnitudes are comparable across arms. Interestingly, this implies that voters were equally as able to infer things like which candidate was better educated and which one had more public office experience by watching the 5 minute get to know you video

as they were after watching 45 minutes of debate. These topics were generally not asked directly, but could plausibly be inferred from the candidate’s manner of speech and physical carriage or confidence.

Moving to policy, row 4 shows that both debates and radio reports enhanced voter knowledge of the candidates’ policy positions. These estimates are much larger in magnitude than the estimate for the get to know you video, which is near zero, and the differences are significant at 99% confidence. The fact that voters learned nothing about policy from the get to know you video, which focused solely on candidate persona, is reassuring for the soundness of the basic research design. Consistent with these estimates, row 5 concerning hypothesis A2 suggests that debates moved voters into better policy alignment with the candidates they selected, while the get to know you video had no impact, a difference that is highly significant (0.074, s.e. 0.033). By contrast to what we see for debates, knowledge of policy positions acquired through the radio report did not translate into better policy alignment. Similarly, only the debates arm had an impact on votes for the debate winner (0.058, s.e. 0.040), which is statistically larger than the result for the radio arm. The fact that radio was equally as effective in building knowledge, but only debates impacted policy preferences and voting choices, suggests a key role for personality in persuading voters to change their behavior. None of the three treatment arms had an impact on crossing party lines, consistent with what we saw earlier for the public screenings, and none of them affected candidate likeability scores.

Overall, while the debate, radio report and get to know you video all affected political knowledge, it is only debates that moved voters to change their voting choices and update their policy views. While this test was not prespecified, we can evaluate whether the treatment effect for debates is larger than the sum of the effects of the radio plus get to know you treatment arms. For policy alignment, the treatment effect for debates is larger than the sum of the other two by 0.114 standard deviation units (s.e. 0.043), and for votes for best, it is larger by 0.098 (s.e. 0.069). Under one sided tests, we can reject the null at 99 and 92 percent confidence, respectively, or at 99 and 85 percent under two sided tests. This pattern of results is consistent with the idea that debates are additive in both charisma and policy/professional information, and that the combination is more powerful than either in isolation.

3.4 Effects of Debate Participation on Elected Members of Parliament (Domain E)

Moving from the election to the behavior of the winning candidates once in office, Table 5 presents treatment effect estimates for the individual outcomes and the four hypothesis-level mean effects indices for this longer term accountability domain. Given the small sample size at this level, standard error estimators that are robust to heteroskedasticity are likely downward biased. To reduce this bias, we thus present standard errors that are the maximum value from conventional ordinary least squares and bias corrected HC₂ estimators in MacKinnon and White (1985), following discussion in Angrist and Pischke (2009). Overall, eight of the eleven treatment effect estimates are positive in sign and five are at least marginally significant. The strong positive effects are concentrated in the latter two hypotheses.

Discussing the hypotheses in order, the first four rows reveal little evidence for treatment effects on the activity level of elected MPs during sittings of Parliament. Specifically, the treatment effect estimates for the percentage of sittings attended and the total number of committees joined are positive but not statistically distinguishable from zero. The estimated effect for total number of public statements made during Parliamentary sittings is negative but not significant, and note the low baseline mean of only 4 statements. These outcomes cover the period from when MPs were inaugurated in December 2012 through the end of 2013, or 57 sittings in total.

The second four rows show no evidence of treatment effects on enhancing topical consistency between the candidate’s first priority sector articulated during the campaign, and their subsequent effort in promoting that sector. We defined the priority sector for each MP based on their pre-election response to the question, “If you had to prioritize one issue in Sierra Leone to receive additional funding in the national budget, what issue would you prioritize?” The modal response was education (44 percent), followed by roads, health and agriculture (each with 15 percent). Treated MPs appear no more likely to have made public statements during a Parliamentary agenda item concerning their preferred sector, although note that only one MP in the entire sample did so. They similarly do not appear more likely to join committees dedicated to that sector, and their constituents are no more likely to report that they focus on that particular sector. Note that we were not able to evaluate consistency in voting in line with pre-stated positions on key national policy issues of interest, as relevant bills have either not yet been introduced (including the gender equity bill) or were passed unanimously (including a freedom of information act).

The next five rows show positive and significant effects of participating in the debates on subsequent constituency engagement. Participating MPs made on average 1.3 (s.e. 0.6)

additional community visits, on a base of 2.9, and held 1.1 (s.e. 0.6) more public meetings, on a base of 1.0. These represent increases of 145 and 210 percent, respectively. Their constituents on average named more sectors in which they viewed the MP as doing “a good job in promoting” that sector in the constituency, and medical staff in clinics were more likely to report that the MP was doing a good job in promoting health. The mean effects index covering all four outcomes is positive and highly significant (0.8 standard deviation units, s.e. 0.3).

The next two rows concern the outcome for which we had the greatest *a priori* interest: the allocation of the constituency facilitation fund (CFF). Recall that this fund of 43.8 M Leones (approximately US\$ 11,000) is given to MPs annually for the development of and transport to their constituencies. MPs are fairly unconstrained in how they spend this money and are not subject to monitoring or reporting requirements. To compile data on how the CFF was spent, we first surveyed each elected MP to generate a detailed itemized list of expenditures and project locations for the first CFF allotment. Our research teams then conducted exhaustive field work to verify these expenditures in the MP’s home constituency, which involved in person visits and physical examination of all purported projects, and multiple interviews with community leaders, clinic staff, teachers and residents of villages where money was reported to have been spent. For the control group, Table 5 shows that only 37 percent of the \$11,000 allotment could be verified as spent on the development of the constituency. The balance was either spent on transport or represents leakage. The treatment effect estimate of 56.1 (s.e. 32.3) suggests that MPs who participated in the debates spent 2.5 times as much on verifiable development expenditures. This effect is significant at 95% confidence. The point estimate corresponds to average gains of roughly six thousand dollars per constituency. Appendix Figure 1 transparently plots the distribution of this outcome by treatment assignment. Comparing the two subplots shows that the positive treatment effect estimate is driven by differences in both tails: there are more zero values amongst the control MPs and more high values amongst the treated MPs. Estimates are robust to dropping the top outlier, which reduces the treatment effect to 48.1 (HC₂ s.e. 29.5, one sided *p*-value 0.06).

The final row estimates the mean effects index across all 11 outcomes in this domain. While this test was not pre-specified, it provides an additional robustness check. The treatment effect estimate suggests that participation in a debate enhanced the post-election performance of MPs by an average of 0.36 standard deviation units (s.e. 0.17), which is significant at above 95% confidence under both one- and two-sided tests.

3.5 Secondary Hypotheses

There were a handful of outcomes that we thought were interesting but less directly related to the debates intervention, so segregated them in the PAP to a more speculative, exploratory category. Appendix Table 2 presents estimates for these outcomes grouped under the three secondary hypotheses of voter turnout, perceived legitimacy of the electoral process, and interest in politics more broadly construed. We find little evidence to support treatment effects for any of these hypotheses.

Specifically, the results for turnout are mixed across voter samples. For the group screening intervention, estimates in the first three rows of Appendix Table 2 reflect negative yet insignificant treatment effects in our exit polls and in the National Electoral Commission (NEC)’s official polling center-level returns. Note that baseline turnout was very high in the control areas, measured at 98.4 percent in our exit poll sample, which is drawn from households in the immediate vicinity of the polling center itself; and 79.1 percent in the NEC returns, which cover voters from the entire catchment area of the polling center. The next three rows concern the individual treatment arms, where we find positive and significant effects for the debate and “get to know you” treatments, and no effect for the radio report. For direct comparability with the estimates in Table 4, recall that these effects are expressed in standard deviation units. To get a better sense of magnitude, note that the treatment effect estimate for the debates treatment is 1.4 percentage points (s.e. 0.69) on a base of 96.1 percent for controls. Since these results do not replicate in the larger polling center level sample, we do not place much weight on them, and conclude that debates exposure if anything had small positive impacts on turnout.

There is no evidence that the debates increased voter confidence that the elections were free and fair, although again baseline confidence was extremely high (91.9 percent for controls). In the bottom of the table, we find some suggestive evidence that exposure to debates spurred voter interest in politics more generally. The mean effects index across the three outcomes of interest is positive and marginally significant. This is driven primarily by a positive effect on voter ability to name the two Presidential candidates, and on voters reporting that they were more likely to discuss politics (although the latter is not statistically significant).

4 Additional Analysis

4.1 Survey Priming

Table 6 turns to the question of how much of these effects can be attributed to the content of the treatment itself as compared to the experience of being surveyed in depth about one’s political views. This is important in light of findings that the act of surveying has nontrivial impacts on behavior (see Zwane et al 2011). The design of our experiment tackles this issue in two ways. First, for the individual-level treatments, we can compare surveyed controls to “pure” controls to capture a survey priming effect. Surveyed controls were given the same detailed survey that accompanied the delivery of the debates and other treatment arms, and then surveyed again in the exit polls. The earlier survey may have primed respondents to seek out information on the outcome variables of interest or increased their salience in the weeks leading up to the Election. By contrast, at the time of treatment implementation the “pure” controls were asked only basic demographic questions, and were not asked about the political outcomes of interest until the exit polls. They thus experienced no prime that would have enhanced their interest in or attention to political information in the weeks leading up to the Election. Analysis in Panel A of Table 6 estimates treatment effects for the surveyed versus pure controls subsample (to capture the survey priming effect), and then separately for the debates treatment versus surveyed controls subsample (to capture the additional treatment effect above and beyond the survey priming effect).

Our second approach uses the group screening sample to capture a survey reinforcing effect by tracking those assigned to treatment with survey versus “pure” treatment across treated and control polling centers. This is thus the converse of the above approach, where we now aim to measure whether being surveyed at the time of treatment facilitates greater comprehension or absorption of the political information conveyed by the debates. Respondents in the treatment plus survey group were given an incentive to attend and also surveyed at the debate screening. Members of the “pure” treatment group were given the attendance incentive but not surveyed at the time of screening. All respondents in the control polling centers are “pure” controls. Analysis in Panel B of Table 6 uses the pooled sample to estimate the coefficient on assignment to a treatment polling center (the pure treatment effect) as well as the interaction term between treatment polling center and assignment to the surveyed group (the additional survey reinforcing effect).

Starting from the right hand side of Table 6 Panel A, estimates in the first row of column 5 suggest that the experience of being surveyed had a small positive effect on political knowledge in the absence of any individual-level treatment. This overall effect is driven by impacts on general political knowledge, where the second row of column 5 suggests that being

surveyed accounts for a 0.099 standard deviation unit (s.e. 0.035) increase in general political knowledge compared to pure controls, significant at 99% confidence. In the left hand side of the same row, column 1 compares those in the debate arm to surveyed controls to reveal a 0.211 standard deviation unit (s.e. 0.042) increase in general political knowledge, which can be attributed to the content of treatment, above and beyond the survey experience. Together, these two estimates suggest that the survey priming effect accounts for one third of the total treatment effect on general political knowledge. Treatment effects on the other two indices of political knowledge (candidate characteristics and policy stances) appear to be driven wholly by the debate content, where treatment effect estimates in column 1 are positive and significant while those in column 5 are not distinguishable from zero. There is weak evidence for a survey priming effect on votes for the debate winner in the sixth row of column 5, however it is noisily estimated and does not hold up in the larger group screening sample of Panel B.

Panel B takes the second approach, where column 1 estimates a “pure” treatment effect from watching the debate without being surveyed. Similar to the above, in the second row there is a 0.238 standard deviation units (s.e. 0.055) treatment effect of watching the debates compared to respondents in control polling centers. In column 5, there is evidence for an additional 0.103 standard deviation unit (s.e. 0.037) effect of being surveyed alongside treatment, suggesting that the survey reinforcing effect similarly accounts for roughly a third of the total effect on general political knowledge. For the other two sub-indices of political knowledge, and for policy alignment and voting for the best performer in the debate, the “pure” treatment effects remain positive and highly significant, and there is no evidence for an additional survey reinforcing effect. There are two marginally significant reinforcing effects on crossing party lines and openness that lack a strong counterpart in the “pure” treatment effect, which would suggest that impacts only arise when debate exposure is followed up with a survey experience. However, these are noisily estimated and do not replicate in the individual-level sample of Panel A.

Overall, these two sets of results suggest that the experience of being surveyed about politics accounts for a third of the total treatment effect on general political knowledge. There is no evidence, however, that survey priming or reinforcing explain much of the other treatment effects described in Section 3. It seems sensible that being surveyed can affect respondent acquisition of straightforward and readily available information, like the eligibility criteria for free healthcare, but not more nuanced information like candidate views on specific policies or which candidate has more education. Given these null results for priming effects on knowledge of candidate attributes and policy stances, it makes sense that we find no evidence for subsequent priming impacts on policy alignment or votes for the candidate who

performed best during the debate.

4.2 Treatment Effect Heterogeneity

Overall, we find little evidence for systematic heterogeneity in treatment effects. Appendix Table 3 estimates heterogeneous effects by respondent sub-groups of gender, age and lack of fluency in Krio (the national *lingua franca* and language of the debates). These specifications use the hypothesis level mean effects index and include all subgroup terms and their interaction with treatment status in a single regression. Across the fifteen estimates of interest, only the negative coefficient on political knowledge for women (-0.076 standard deviation units, s.e. 0.021) is significant at conventional levels (under two sided tests). In terms of magnitude, this suggests that women acquired only 75% as much political knowledge from the debates when compared to men.

Our results also do not appear to be driven by large effects in any particular constituency. As an example, the treatment effect estimate on voting for the best candidates is robust to excluding each constituency one by one.

Regarding dissipation of treatment effects over time, we find suggestive evidence for an immediate drop off in political knowledge gains in the few days after treatment, but no evidence that this dissipation intensifies with the lag between treatment exposure and the exit poll. Confining our attention to the treatment group only, voter knowledge doubled from the before- to after-screening survey: voters on average could correctly answer 24 percent of political knowledge questions at baseline, which jumped to 46 percent immediately after watching the group screening. By the time of the exit poll, this percentage had fallen to 40, implying that roughly one third of the initial knowledge gains had dissipated. This implies persistent gains of 16 percentage points, or a 66 percent increase on baseline knowledge. Similar estimates obtain for those in the individual-level debate treatment arm, save voters in these areas began from a higher base: 29 percent correct answers on average in the before survey, increasing to 46 percent in the after survey, and falling to 42 percent in the exit poll. Bringing in the control group, we next estimate whether this attenuation varies with the time lag between the screening and the exit poll, which varies from 6 to 35 days. Note that this time variation is not random, so estimates rely on the assumption that factors determining field team deployment (e.g. remoteness) are orthogonal to voter responsiveness to treatment. Here we find no systematic evidence for heterogeneity over time: treatment effects for those treated far from the election, e.g. 30 days or more, are very similar to estimates for

treated close to the election, e.g. within 10 days of the exit poll. Our interpretation is that some knowledge gains dissipate quickly after exposure, while the remaining gains

observed a week later persist for several additional weeks.

Some additional heterogeneity analyses remain to be compiled.

4.3 Social Mobilization: Individual versus Group Exposure

This section considers how the delivery of the debate content—via group screening versus individual private viewing—affects the impact it has on voter behavior. Since many aspects of the experience differ across the two delivery modes, we will not be able to pin down exact mechanisms, but can speculate as to how salient differences might drive divergence in treatment effect intensity. First note that the content of the debate films was exactly the same under the two conditions. The most pronounced difference in delivery is that the screenings involved large public gatherings of a couple hundred people, while the individual treatment arm involved respondents watching the debate alone on a tablet device. The PAP thus classifies this comparison of the two delivery modes as a test of social mobilization effects. Lab experiments have shown that exposure to the reactions of audience members—either real or fabricated—can have significant effects on evaluations of debate performance and candidate attributes (Fein et al. 2007, Davis et al. 2011). Note, however, that other aspects differ across the two modes: individual treatments were administered in larger polling centers (as measured by total registered voters); only the group screenings attracted a campaign spending response; and the implementation procedures varied, where specifically group screenings played music before the debates, played the debates twice, had simultaneous translation into the relevant local language, and occurred later in the day.

Table 7 presents the cleanest comparison of the two delivery mechanisms by limiting the group screening estimates to the 8 constituencies where the individual treatments were also implemented, and restricting the individual estimates to a comparison of the debates treatment arm to the pure control group (dropping respondents from the surveyed control and lab experiment control arms). First note that the qualitative pattern of effects for the two delivery modes on these comparable subsamples is the same: strong positive treatment effects on political knowledge, including effects for all three sub-indices, on policy alignment, and on votes for the best performer; and no evidence of treatment effects on crossing party lines or voter openness. Second note that the treatment effect for the group screening is larger in magnitude than that of the individual viewing everywhere save on votes for the best performer, where it is equal. This difference is more pronounced when we scale up the intention to treat effects for the group level screening to estimate treatment on the treated effects (Column 3), which are more directly comparable to the individual level treatments where compliance with treatment assignment was nearly perfect. Notice that the difference

in magnitude is most pronounced for knowledge of candidate characteristics, knowledge of candidate policy positions, and moving into policy alignment. These differences are consistent with the idea that watching the films in a group setting facilitated greater discussion amongst voters that clarified and reinforced the information about candidates and policy conveyed by the debates. The fact that the impact on general political knowledge is comparable across the two modes suggests that basic differences in comprehension (attributable to the lack of local language translation in the individual-level arms) cannot fully explain the divergence in magnitude of effect.

5 Conclusion

These experiments suggest that voters acquire significant political knowledge from watching candidate debates, knowledge that persists over a number of weeks, and importantly, influences their vote choice on Election Day. By equipping voters with knowledge that changes their voting behavior, debates further attract greater campaign investment by participating candidates. This spending response is consistent with debate exposure making vote margins appear narrower or more uncertain *ex ante*, even in areas where it was revealed *ex post* that debates favored the more popular candidate. Debates convey comprehensive information about candidates—including charisma, professional qualifications and policy stances—and the combination of factors appears more powerful than each in isolation. Over the longer run, participation in debates appears to enhance accountability pressure of elected officials, increasing their subsequent engagement with constituents and expenditure on development projects. The finding that debates can enhance accountability even in relatively uncompetitive areas, where direct electoral pressure is limited, is important.

From a policy perspective, this project demonstrates that interparty debates between candidates are logistically feasible to host and disseminate; and can create large, persistent gains in voters’ political knowledge. Commissioners of the National Electoral Commission of Sierra Leone, who have been consulted on this project from its inception, responded positively to our early dissemination events and expressed interest in taking debates to scale in the next elections. In considering the costs and benefits of scaling up implementation, fixed video production costs for the debates themselves were modest in this setting: roughly five thousand dollars per constituency to assemble the candidates, host, film and edit the debates. The point estimate on increased development expenditure associated with debate participation is large enough to fully cover this cost. In terms of marginal dissemination costs, the mobile cinema was a relatively resource intensive way to publicize the debates. In settings where mass media penetration is higher, televising the debates would be an obvious

alternative dissemination mechanism with lower marginal costs. Where television broadcast or ownership is limited, radio becomes the obvious choice. While the individual treatment arms suggest that video is more effective than audio alone, the radio report we tested was rather dry. One could imagine a livelier radio program that captures a real time debate between candidates in the recording studio that might come closer to the impacts of the film screening. This kind of radio counterpart could reach large voting audiences at negligible marginal cost. Moreover, assuming that accountability effects are increasing in the share of the public exposed to debates, increasing the scale of dissemination would likely enhance the observed accountability effects and implicit cost recovery via development spending.

References

- Abramowitz, Alan I. "Impact of a Presidential Debate on Voter Rationality," *American Journal of Political Science*, 22:3(1978): 680-690.
- Angrist, Joshua D. and Jorn-Steffen Pischke. *Mostly Harmless Econometrics*. Princeton: Princeton University Press, 2009.
- Ansolabehere, Stephen and James M. Snyder Jr. "The Incumbency Advantage in U.S. Elections: An Analysis of State and Federal Offices, 1942-2000." *Election Law Journal*, 1:3(2002), 315-338.
- Banerjee, Abhijit V., Selvan Kumar, Rohini Pande and Felix Su. "Do Informed Voters Make Better Choices? Experimental Evidence from Urban India." manuscript, 2011.
- Bardhan, Pranab and Dilip Mookherjee, "Determinants of Redistributive Politics: An Empirical Analysis of Land Reforms in West Bengal, India," *American Economic Review*, 100:4(2010), 1572-1600.
- Benjamin, Daniel J. and Jesse M. Shapiro. "Thin-slice Forecasts of Gubernatorial Elections." *Review of Economics and Statistics*, 91:3(2009): 523–536.
- Casey, Katherine. "Crossing Party Lines: The Effects of Information on Redistributive Politics." *American Economic Review* (forthcoming).
- Casey, Katherine, Rachel Glennerster and Edward Miguel. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." *Quarterly Journal of Economics*, 127:4(2012): 1755-1812.
- Cranes-Wrone, Brandice, Michael C. Herron and Kenneth W. Shotts. "Leadership and Pandering: A Theory of Executive Policymaking." *American Journal of Political Science*, 45:3(2001):532-550.
- Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot and Philippe Zamora. "Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment." *Quarterly Journal of Economics* (2013), 531–580.
- Davis, Colin J., Jeffrey S. Bowers and Amina Memon. "Social Influence in Televised Election Debates: A Potential Distortion of Democracy." *PLOS One*, 6:3(2011), e18154.
- Dixit, Avinash, and John Londregan, "The Determinants of Success of Special Interests in Redistributive Politics," *The Journal of Politics*, 58(1996), 1132-1155.
- Dixit, Avinash, and John Londregan, "Ideology, Tactics, and Efficiency in Redistributive Politics," *Quarterly Journal of Economics*, (1998), 497-529.
- Downs, Anthony (1957). *An Economic Theory of Democracy*. New York: Harper.
- Druckman, James N. "The Power of Television Images: The First Kennedy-Nixon Debate Revisited." *Journal of Politics*, 65:2(2003): 559–571.

- Fein, Steven, George R. Goethals and Matthew B. Kugler. "Social Influence on Political Judgments: The Case of Presidential Debates." *Political Psychology*, 28:2(2007), 165-192.
- Ferraz, Claudio and Frederico Finan. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics*, 123:2(2008): 703-745.
- Fridkin, Kim L., Patrick J. Kenney, Sarah Allen Gershon, Karen Shafer and Gian Serignese Woodall. "Capturing the Power of a Campaign Event: The 2004 Presidential Debate in Tempe." *Journal of Politics*, 69:3(2007), 770-785.
- Fujiwara, Thomas, and Leonard Wantchekon. "Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin." *American Economic Journal: Applied Economics*, 5:4(2013): 241-55.
- Gerber, Alan S., James G. Gimpel, Donald P. Green, and Daron R. Shaw. "How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment." *American Political Science Review*, 105:1(2011): 135-150.
- Hellweg, Susan A., Michael Pfau and Steven R. Brydon. *Televised Presidential Debates: Advocacy in Contemporary America*. Praeger, 1992.
- Humphreys, Macartan and Jeremy M. Weinstein. "Policing Politicians: Citizen Empowerment and Political Accountability in Uganda Preliminary Analysis." manuscript, 2012.
- Jamieson, Kathleen Hall and David S. Birdsell. *Presidential Debates: The Challenge of Creating an Informed Electorate*. Oxford University Press, 1990.
- Kendall, Chad, Tommaso Nannicini and Francesco Trebbi. "How Do Voters Respond to Information? Evidence from a Randomized Campaign." *American Economic Review*, 105:1(2015), 322-353.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75:1(2007), 83-119.
- Lenz, Gabriel S. "Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis." *American Journal of Political Science*, 53:4(2009), 821-837.
- Liessem, Verena, and Hans Gersbach. "Incentive Contracts and Elections for Politicians with Multi-Task Problems." Available at SSRN 243518 (2003).
- MacKinnon, James G. and Halbert White. "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics*, 29(1985), 305-325.
- McKinnon, Lori Melton, John C. Tedesco and Lynda Lee Kaid. "The Third 1992 Presidential Debate: Channel and Commentary Effects." *Argumentation and*

- Advocacy, 30:2(1993), 106-118.
- Mullainathan, Sendhil, Ebonya Washington and Julia R. Azari. "The impact of electoral debate on public opinions: an experimental investigation of the 2005 New York City mayoral election," in Shapiro, Ian, Susan C. Stokes, Elisabeth Jean Wood, and Alexander S. Kirshner, eds. Political representation. Cambridge University Press, 2010.
- Prat, Andrea. "The Wrong Kind of Transparency." *American Economic Review*, 95:3(2005): 862-877.
- Prior, Markus. "Who watches presidential debates? Measurement problems in campaign effects." *Public Opinion Quarterly*, 76:2(2012): 350-363.
- Shear, Michael D. "After Debate, a Torrent of Criticism for Obama," *New York Times*, October 5, 2012.
- Todorov, Alexander, Anesu N. Mandisodza, Amir Goren, and Crystal C. Hall. "Inferences of competence from faces predict election outcomes." *Science*, 308:5728(2005): 1623-1626.
- Wald, Kenneth D. and Michael B. Lupfer. "The Presidential Debate as a Civics Lesson." *Public Opinion Quarterly*, 42:3(1978), 342-353.
- Wantchekon, Leonard. "Clientelism and Voting Behavior: Evidence from A Field Experiment In Benin." *World Politics*, 55(2003): 399-422.
- Wantchekon, Leonard, Gabriel Lopez-Moctezuma, Thomas Fujiwara, Cecilia Pe Lero and Daniel Rubenson. "Policy Deliberation and Voting Behavior: A Campaign Experiment in the Philippines." manuscript, Princeton University (2014).
- Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer et al. "Being surveyed can change later behavior and related parameter estimates." *Proceedings of the National Academy of Sciences* 108:5(2011): 1821-1826.

Table 1: Domain A - Treatment Effects of Polling Center Debate Screenings on Voters

Mean Effects Indices by Hypothesis	Treatment effect (standard error)	Per comparison p-value (one sided)	FWER adjusted p-value (one sided)
	(1)	(2)	(3)
A1. Exposure to debates increases political knowledge (20 outcomes)	0.281 (0.028)	0.000**	0.000**
A2. Exposure to debates increases policy alignment (3 outcomes)	0.106 (0.035)	0.002**	0.009**
A3. Exposure to debates increases vote shares for the candidate that performed the best in the debates (2 outcomes)	0.086 (0.043)	0.023*	0.079+
A4. Exposure to debates increases the willingness to vote across party lines (3 outcomes)	-0.018 (0.032)	0.718	0.710
A5. Exposure to debates enhances voter openness to other parties (5 outcomes)	0.091 (0.048)	0.028*	0.079+

Notes: i) significance levels indicated by + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$; ii) robust standard errors clustered by polling center; iii) p-values are based on one-sided tests in the direction of the hypothesis statement as pre-specified in our pre-analysis plan (PAP); iv) all specifications include stratification bins for the polling center (number of registered voters and distance to next nearest), respondent (youth status and gender) and constituency fixed effects; v) specifications further include the full set of control variables (gender, age, years of schooling, polygamous marital status, farming occupation and radio ownership); vi) treatment effects are on the hypothesis-level mean effects index constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units; vii) adjustments to control familywise error rate (FWER) computed following Westfall and Young (1993) and Anderson (2008); and viii) data source is the exit poll survey.

Table 2: Domain A Treatment Effects on All Individual Outcomes

Individual outcomes by hypothesis	Control mean (1)	Treatment effect (2)	Standard error (3)	Naïve p-value (4)	FDR q-value (5)	N (6)
Panel A1. Political knowledge increases						
i. Knows amount of the constituency facilitation fund	0.034	0.140	0.018	0.000**	0.001**	5,400
i. Knows who is entitled to free healthcare (FHC)	0.706	0.056	0.033	0.044*	0.040*	5,399
i. Knows the gender equity bill (GEB) is 30%	0.352	0.012	0.030	0.346	0.237	5,398
i. Knows MP job responsibilities (out of 3)	0.555	0.217	0.070	0.001**	0.003**	5,400
ii. Candidate choice driven by individual characteristic	0.209	-0.010	0.028	0.638	0.350	5,229
ii. Knows which candidates had been an MP before	0.490	0.111	0.032	0.000**	0.002**	5,400
ii. Knows which candidate was most educated	0.243	0.159	0.044	0.000**	0.001**	3,097
ii. Knows candidate with most public office experience	0.319	0.076	0.037	0.021*	0.087+	2,972
ii. Knows APC candidate's name	0.442	0.181	0.034	0.000**	0.001**	5,058
ii. Knows PMDC candidate's name	0.115	0.105	0.031	0.000**	0.002**	3,291
ii. Knows SLPP candidate's name	0.395	0.168	0.031	0.000**	0.001**	5,400
iii. Knows APC candidate's first priority issue	0.190	0.088	0.030	0.002**	0.005**	5,057
iii. Knows PMDC candidate's first priority issue	0.099	0.064	0.026	0.008**	0.012*	3,288
iii. Knows SLPP candidate's first priority issue	0.142	0.149	0.028	0.000**	0.001**	5,398
iii. Knows APC candidate's view of FHC	0.252	0.197	0.035	0.000**	0.001**	4,579
iii. Knows PMDC candidate's view of FHC	0.119	0.007	0.037	0.421	0.259	2,812
iii. Knows SLPP candidate's view of FHC	0.123	0.072	0.029	0.007**	0.011*	4,921
iii. Knows APC candidate's position on GEB	0.285	0.096	0.035	0.003**	0.007**	5,058
iii. Knows PMDC candidate's position on GEB	0.244	0.209	0.053	0.000**	0.001**	3,291
iii. Knows SLPP candidate's position on GEB	0.331	0.155	0.038	0.000**	0.001**	5,400
Panel A2. Policy alignment increases						
Voter's view on FHC matches that of chosen candidate	0.394	0.092	0.035	0.004**	0.008**	4,727
Voter's view on GEB matches that of chosen candidate	0.613	-0.024	0.024	0.843	0.432	5,160
Voter's priority issue matches that of chosen candidate	0.425	0.090	0.031	0.002**	0.005**	5,160
Panel A3. Votes for best performing candidate in the debate increase						
Voted for debate winner, as judged by audience	0.803	0.049	0.021	0.012*	0.015*	5,219
Voted for debate winner, as judged by expert panel	0.712	0.011	0.022	0.312	0.218	5,219
Panel A4. Votes across ethnic-party lines increase						
Voted across ethnic-party lines	0.107	-0.012	0.013	0.811	0.432	4,569
Voted for a different party for MP than did in 2007	0.163	0.004	0.019	0.425	0.259	4,405
Split ticket across parties for MP vs President	0.058	-0.009	0.010	0.833	0.432	5,212
Panel A5. Voter openness to candidates increases						
Voter likeability rank for APC candidate (10 point scale)	6.523	0.161	0.261	0.269	0.212	5,073
Voter likeability rank for own party's candidate	7.971	0.245	0.237	0.151	0.120	5,160
Voter likeability rank for PMDC candidate	2.369	0.579	0.273	0.018*	0.020*	3,299
Voter likeability rank for rival party's candidate	3.395	0.117	0.226	0.303	0.218	4,906
Voter likeability rank for SLPP candidate	4.952	0.229	0.230	0.160	0.123	5,414
Panel A3 Alternate: NEC Official Returns						
Vote share of debate winner, as judged by audience	0.711	0.035	0.017	0.033*		206
Vote share of debate winner, as judged by expert panel	0.617	0.035	0.016	0.027*		206

Notes: i) significance levels + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ based on one-sided tests in direction pre-specified in PAP; ii) robust standard errors clustered by polling center; iv) specifications include stratification bins for the polling center (number of registered voters and distance to next nearest), respondent (youth status and gender) and constituency fixed effects; v) additional controls vary by hypothesis from the set (gender, age, years of schooling, polygamous marital status, farming occupation and radio ownership); vi) adjustments to control false discovery rate (FDR) computed following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008) across all 48 primary and secondary individual outcomes in domain A; viii) data source is the exit poll survey in panels A1-A5 and the National Electoral Commission polling center-level returns in A3 Alt; and ix) NEC returns exclude constituency 15 where the SLPP candidate was disqualified immediately before the election.

Table 3: Domain B - Treatment Effects of Polling Center Screenings on Candidates

Outcome	Control mean	Treatment effect	Standard error	Naïve p-value (2 sided)	FDR q-value	N
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Hypothesis A. Mean Effects Index (all 9 outcomes)</i>	0.000	0.103	0.039	0.008**		5,399
Received any gift from the APC candidate	0.160	0.011	0.027	0.686	0.666	5,055
Received any gift from the PMDC candidate	0.007	0.013	0.006	0.027*	0.089+	3,220
Received any gift from the SLPP candidate	0.089	0.007	0.020	0.725	0.666	5,397
Value of gift received from APC candidate (in log(value+1))	0.412	0.121	0.098	0.217	0.364	4,989
Value of gift received from PMDC candidate (in log(value+1))	0.014	0.034	0.014	0.016*	0.089+	3,213
Value of gift received from SLPP candidate (in log(value+1))	0.210	0.077	0.063	0.222	0.364	5,347
Voter report of number of APC candidate visits to village	1.292	0.147	0.137	0.285	0.398	5,056
Voter report of number of PMDC candidate visits to village	0.353	0.219	0.093	0.019*	0.089+	3,291
Voter report of number of SLPP candidate visits to village	1.273	0.070	0.186	0.709	0.666	5,399

Notes: i) significance levels + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$; ii) robust standard errors clustered by polling center; iii) specifications include stratification bins for the polling center (number of registered voters and distance to next nearest), respondent (youth status and gender) and constituency fixed effects; iv) additional controls determined by analysis of control group data and include gender, age, years of schooling, and radio ownership; v) mean effects index constructed following Kling, Liebman and Katz 2007 and is expressed in standard deviation units; vi) adjustments to control false discovery rate (FDR) computed following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008); and vii) data source is the exit poll survey.

Table 4: Domain D - Causal Mechanisms Explored through Relative Treatment Effects Across Individual Treatment Arms

Hypothesis		Mean Effects Index												
	Debate	Naive	Get to Know You	Radio Report	Debate vs. GTKY	Debate vs. Radio	Radio vs. GTKY							
	Treatment effect (std error)	Naive p value 1 sided (1)	Treatment effect (std error)	Naive p value 1 sided (2)	Treatment effect (std error)	Naive p value 1 sided (3)	Treatment effect (std error)	Naive p value 1 sided (4)	Treatment effect (std error)	Naive p value 1 sided (5)	Treatment effect (std error)	Naive p value 1 sided (6)	Treatment effect (std error)	Naive p value 1 sided (7)
A1. Political knowledge	0.109** (0.021)	0.000 (2)	0.041** (0.016)	0.006 (4)	0.095** (0.018)	0.000 (6)	0.068** (0.022)	0.002 (8)	0.014 (0.018)	0.425 (10)	0.053* (0.022)	0.016 (12)		
i. General Knowledge	0.175** (0.040)	0.000 (2)	0.095** (0.035)	0.005 (4)	0.160** (0.045)	0.000 (6)	0.079+ (0.043)	0.066 (0.197)	0.014 (0.034)	0.674 (0.736)	0.065 (0.050)	0.192 (0.370)		
ii. Candidate Characteristics	0.049** (0.019)	0.006 (2)	0.068** (0.025)	0.005 (4)	0.042* (0.020)	0.021 (6)	-0.019 (0.026)	0.455 (0.521)	0.007 (0.026)	0.793 (0.819)	-0.026 (0.032)	0.411 (0.521)		
iii. Policy Stances	0.127** (0.031)	0.000 (2)	-0.003 (0.017)	0.575 (4)	0.106** (0.023)	0.000 (6)	0.130** (0.028)	0.000 (0.001)	0.020 (0.026)	0.434 (0.521)	0.110** (0.026)	0.000 (0.001)		
A2. Policy Alignment	0.081** (0.029)	0.004 (2)	0.007 (0.027)	0.395 (4)	-0.040 (0.024)	0.945 (6)	0.074* (0.033)	0.025 (0.101)	0.121** (0.032)	0.000 (0.002)	-0.047+ (0.027)	0.083 (0.199)		
A3. Vote for best	0.058+ (0.040)	0.077 (2)	0.006 (0.037)	0.440 (4)	-0.046 (0.043)	0.851 (6)	0.052 (0.045)	0.241 (0.386)	0.104* (0.052)	0.046 (0.159)	-0.051 (0.040)	0.203 (0.370)		
A4. Cross party lines	-0.030 (0.035)	0.802 (2)	0.004 (0.031)	0.453 (4)	0.058 (0.045)	0.103 (6)	-0.033 (0.044)	0.447 (0.521)	-0.088+ (0.050)	0.076 (0.199)	0.055 (0.042)	0.195 (0.370)		
A5. Openness	0.006 (0.023)	0.395 (2)	-0.022 (0.025)	0.812 (4)	0.014 (0.030)	0.322 (6)	0.029 (0.034)	0.403 (0.521)	-0.008 (0.033)	0.818 (0.819)	0.036 (0.029)	0.215 (0.370)		
Number of observations	1,698		1,695		1,695									

Notes: i) significance levels, corresponding to one sided per comparison p values, + p < 0.10, * p < 0.05, ** p < 0.01; ii) robust standard errors clustered by polling center; iii) specifications include stratification bins for the household (gender and age composition), polling center (number of registered voters and distance to next nearest) and constituency fixed effects; iv) mean effects index constructed following Kling, Liebman and Katz 2007 and is expressed in standard deviation units; v) Columns 1 to 6 run one-sided tests in the (prespecified) direction of the hypothesis statement, while Columns 7 to 12 run two-sided tests; vi) adjustments to control false discovery rate (FDR) computed following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008) across all 24 tests run; and vii) data source is the individual treatment sample exit poll survey.

Table 5: Domain E - Treatment Effects of Debate Participation on Accountability

Outcomes by hypothesis	Control mean	Treatment effect	Standard error max(OLS, HC ₂)	Naïve p-value (1 sided)	N
	(1)	(2)	(3)	(4)	(5)
<i>Hypothesis E1. Activity in Parliament, mean effects index</i>	0.000	0.129	0.320	0.345	28
Percent of 2012-13 sittings attended (out of 57 sittings in total)	76.692	3.371	3.003	0.137	28
Total number of public comments in Parliamentary sittings 2012-13	4.286	-1.569	2.224	0.780	28
Committee membership (total number)	3.929	0.524	0.625	0.186	28
<i>Hypothesis E2. Consistency with pre-election promises, mean effects index</i>	0.000	-0.219	0.226	0.829	28
Total public comments in priority sector agenda items	0.154	-0.170	0.166	0.842	27
Membership in priority sector committee	0.231	0.201	0.187	0.147	27
Constituent assessment of focus on priority sector	0.571	-0.343	0.150	0.984	27
<i>Hypothesis E3. Constituency engagement, mean effects index</i>	0.000	0.779	0.307	0.009**	28
Total number of constituent visits	2.915	1.316	0.619	0.022*	28
Total number of public meetings held with constituents	1.018	1.089	0.606	0.043*	28
Total number of sectors constituents assess good performance	1.417	0.882	0.476	0.039*	28
Health clinic staff reported good performance in health	0.202	0.187	0.135	0.089+	28
<i>Hypothesis E4. CFF spending, mean effects index</i>	0.000	1.139	0.623	0.041*	28
Development spending verified in the field (as % of 2012 CFF allotment)	37.743	56.081	32.271	0.049*	27
<i>Domain E. All 11 outcomes, mean effects index</i>	0.000	0.362	0.168	0.021*	28

Notes: i) significance levels + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$; ii) the standard error (and associated p value) presented is the maximum value from conventional OLS and bias corrected HC₂ estimators in MacKinnon and White (1985), following discussion in Angrist and Pischke (2009); iii) specifications include stratification bins for the constituency (3 bins of ethnic-party bias), MP gender and an indicator for whether the MP held an elected position in the past; iv) mean effects index constructed following Kling, Liebman and Katz 2007 and is expressed in standard deviation units; and v) missing values for hypothesis E2 concern one control MP who did not provide a pre-election priority sector and for hypothesis E4 concern one treated MP who did not take office until December, 2013 (one year after the election) and thus did not receive the 2012 CFF.

Table 6: Survey Priming and Reinforcing Effects

Panel A: Survey priming effects across individual treatment arms

Mean Effects Index by Hypothesis	Treatment effect beyond survey:				Survey priming effect:			
	Debates vs Surveyed controls				Surveyed controls vs. Pure controls			
	Treatment effect (1)	Standard error (2)	Naïve p-value (3)	N (4)	Treatment effect (5)	Standard error (6)	Naïve p-value (7)	N (8)
A1. Political knowledge	0.118**	0.023	0.000	875	0.026+	0.015	0.051	875
i. General Knowledge	0.211**	0.042	0.000	875	0.099**	0.035	0.004	875
ii. Candidate Characteristics	0.057**	0.023	0.008	875	-0.021	0.019	0.858	875
iii. Policy Stances	0.125**	0.033	0.000	875	0.030	0.024	0.114	875
A2. Policy Alignment	0.079**	0.032	0.010	933	0.012	0.038	0.372	935
A3. Vote for best	0.025	0.053	0.317	933	0.079+	0.053	0.070	935
A4. Cross party lines	-0.010	0.039	0.605	933	-0.045	0.046	0.834	935
A5. Openness	-0.003	0.025	0.546	933	0.006	0.028	0.411	935

Panel B: Survey reinforcing effects across group screening arms

Mean Effects Index by Hypothesis	"Pure" treatment effect:				Survey reinforcing effect:			
	Debates without survey vs. controls				Interaction term for debates with survey			
	Treatment effect (1)	Standard error (2)	Naïve p-value (3)	N (8)	Treatment effect (5)	Standard error (6)	Naïve p-value (7)	N (8)
A1. Political knowledge	0.263**	0.031	0.000	5,415	0.028+	0.020	0.079	5,415
i. General Knowledge	0.238**	0.055	0.000	5,415	0.103**	0.037	0.003	5,415
ii. Candidate Characteristics	0.247**	0.036	0.000	5,415	0.010	0.025	0.340	5,415
iii. Policy Stances	0.286**	0.039	0.000	5,415	0.009	0.022	0.344	5,415
A2. Policy Alignment	0.105**	0.041	0.005	5,415	-0.001	0.030	0.510	5,415
A3. Vote for best	0.088*	0.047	0.033	5,415	-0.015	0.036	0.664	5,415
A4. Cross party lines	-0.054	0.037	0.925	5,415	0.056+	0.034	0.052	5,415
A5. Openness	0.064	0.050	0.103	5,415	0.042+	0.026	0.052	5,415

Notes: i) significance levels, corresponding to one sided per comparison p values, + p < 0.10, * p < 0.05, ** p < 0.01; ii) robust standard errors clustered by polling center; iii) specifications include stratification bins for the relevant randomization procedure and constituency fixed effects; iv) mean effects index constructed following Kling, Liebman and Katz 2007 and expressed in standard deviation units; and v) mean effects indices are standardized with respect to the pure control group for all of panel A.

Table 7: Social Mobilization Effect - Group vs Individual Delivery of Debates

Mean Effects Index by Hypothesis	Group Screening			Individual Viewing	
	Treatment effect (ITT) (std error)	TOT scaling	N	Treatment effect (\approx TOT) (std error)	N
	(1)	(2)	(3)	(4)	(5)
A1. Political knowledge	0.324** (0.034)	0.416	3,507	0.140** (0.023)	708
i. General political knowledge	0.371** (0.034)	0.475	3,507	0.304** (0.053)	708
ii. Candidate characteristics	0.268** (0.039)	0.344	3,507	0.031+ (0.023)	708
iii. Policy stances	0.348** (0.049)	0.446	3,507	0.152** (0.037)	708
A2. Policy alignment	0.129** (0.047)	0.165	3,514	0.087* (0.039)	748
A3. Votes for best performer in the debate	0.102+ (0.068)	0.130	3,510	0.102* (0.051)	748
A4. Cross party lines	-0.022 (0.042)	-0.028	3,511	-0.052 (0.050)	748
A5. Openness	0.059 (0.065)	0.075	3,514	-0.002 (0.031)	748

Notes: i) significance levels, corresponding to one sided per comparison p values, + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$; ii) robust standard errors clustered by polling center; iii) specifications include stratification bins for the randomization procedure and constituency fixed effects; iv) additional controls vary by dataset and hypothesis from the set (gender, age, years of schooling, polygamous marital status, farming occupation and radio ownership); v) mean effects index constructed following Kling, Liebman and Katz 2007 and expressed in standard deviation units; vi) the group screening sample is limited to the 8 constituencies where the individual-level treatments were also administered; vii) the individual level estimates are limited to the debates treatment arm versus the "pure" control group.

Online Appendix

for “Debates: The Effects of Voter Knowledge Initiatives in Sierra Leone”

by Bidwell, Casey and Glennerster

Draft: March 2015

SIERRA LEONE 2012 ELECTIONS PROJECT

PRE-ANALYSIS PLAN: POLLING CENTER LEVEL INTERVENTIONS

PIs: Kelly Bidwell (IPA), Katherine Casey (Stanford GSB) and Rachel Glennerster (JPAL MIT)

Revised Plan: 12 September 2013

This study examines the impact of providing citizens with information about Parliamentary candidates via structured inter-party debates in the lead up to the Sierra Leone November 2012 Elections. Randomization and treatments were conducted on multiple levels: constituency, polling center and individual (details on sampling and randomization are available in the project's [AEA trial registry https://www.socialscienceregistry.org/trials/26](https://www.socialscienceregistry.org/trials/26)). This pre-analysis plan governs the analysis of the polling-center level treatment only. The first version of this plan was written and registered with the Abdul Latif Jameel Poverty Action Lab on 20 November 2012, before fieldwork for the exit poll, which is the primary source of data for this analysis, was completed. This revised plan incorporates learning from the following steps that we have taken since registering the initial plan, namely we: (i) analyzed the expert panel scoring of debates and the before/after debate surveys; (ii) registered a separate plan for the individual-level treatments; (iii) analyzed treatment effects for the individual-level treatments; and (iv) examined the distribution of outcomes for the control group polling centers in the exit poll data. We are now registering an update to the initial document reflecting learning from steps 1 to 4; before we analyze treatment effects at the polling-center level in the exit poll. Planned future steps include: i) lodging an update governing the analysis of the electoral returns data before completing that portion of the analysis (which depends on two additional datasets that have not yet been cleaned); and ii) lodging an update governing the analysis of constituency-level effects (as this data collection effort remains ongoing).

1. Background

Our NGO partner, Search for Common Ground, hosted and filmed debates between Parliamentary candidates in 14 constituencies. We randomly selected these constituencies from what we estimated would be the 28 most competitive constituencies, stratifying on the strength of the ethnic bias favoring one party over the other. Within constituencies, polling centers that were sufficiently small (fewer than ~900 registered voters) and far apart from their nearest neighbor (at least ~ one mile) were randomly assigned to treatment and control groups.

Treatment at the polling center level consisted of an evening showing of a video tape of the relevant debate projected at a convenient public place, usually the polling center itself, in the weeks leading up to the Election. Typical protocol for these screenings was as follows: host polling center and satellite communities were notified in advance and invited to attend the screening; 25 randomly selected residents (using data from an earlier listing exercise) were provided a small incentive (10 Maggi spice cubes for cooking) to attend the screenings; the video was played once in a pause and play format that inserted translation into the relevant local language after each question; the video was played a second time with or without translation; and a secondary screening was held in the largest accessible satellite village earlier in the day, in most cases without translation.

We hypothesize that this video screening intervention may have treatment effects on three different sets of actors: voters, candidates and centralized parties, which we will treat as distinct domains. For each set,

Comment [KC1]: For transparency, we have tracked the changes we made to the original PC-level PAP lodged on 20 Nov 2012 and included explanatory comments for the more substantive revisions.

Deleted: 20 November 2012

Deleted: "Sampling Procedures" document

Deleted: It

Comment [KC2]: We changed the planned order of our analysis to complete more of the exploratory work before embarking on the PC-level analysis.

Deleted: (where the current estimated completion date is 22 November 2012). This document is the first installment in a planned sequence of registry and data analysis, where we will next:

Deleted: s

Deleted: and constituency-level

Deleted: ii

Deleted: ; (iv) analyze the expert panel scoring of debates and the before/after debate surveys; (v)

Deleted: this

Deleted: 2

Deleted: and then (vi)

Deleted: and voting returns data

we lay out a series of hypotheses regarding the likely areas of impact with corresponding outcome measures below. We will provide treatment effects with unadjusted (or per comparison) p-values for all outcomes specified in this document for all domains. We will also compute mean effects indices by hypothesis (and sub-hypothesis for H3) and correct for multiple inference across outcomes within a hypothesis. Since we have multiple hypotheses regarding voters, we will further make adjustment at the hypothesis-level in domain A (as specified below) but will not make adjustments across domains.

2. Domain A: Effects of PC-level Debates on Voters

This domain explores the effects of polling center debate screenings on voter knowledge, behavior and vote choice.

A. Econometric specifications

Analysis of treatment effects will take the form of:

$$Y_{ipc} = \beta_0 + \delta T_{pc} + X'_{ipc}\Pi + Z'_{pc}\Gamma + W'_{ipc}\Psi + c_p + \varepsilon_{ipc} \quad (1)$$

where outcome Y (i.e. vote choice) is measured for individual i registered in polling center p within Parliamentary constituency c ; T is an indicator variable equal to one if the polling center received the debate video screening treatment; X is a vector of indicator variables that denote the stratification bin from which exit poll respondents were drawn (where the bins were constructed by age and gender); Z is a vector of indicator variables that denote the stratification bin from which the polling center was drawn (where the bins were constructed by number of registered voters and distance to nearest neighboring center); W is a set of additional control variables determined from analysis of the control group data and will vary by hypothesis with an eye toward identifying individual characteristics that do not vary with treatment and that help explain variation in a particular outcome (see algorithm below); c is a set of constituency-specific fixed effects (the level of debate and candidates); and ε is an idiosyncratic error term clustered at the polling center level. Our main specification includes the full set of controls (X , Z and W); we will also show results for the sparser specification that includes only the stratification variables as controls (X and Z only) as a robustness check. We will determine W as any subset of {gender, age, frequency of discussing politics, education, marital status, occupation, radio ownership} that predicts outcomes for the control group with at least 95% confidence. The coefficient of interest is δ , the average treatment effect. Unless otherwise stated, all tests will be one-sided in the direction indicated below. The primary source of data is the individual-level exit polls.

We will test for heterogeneous treatment effects at the level of constituency, candidate and voter, adjusting for multiple inference within each level (i.e. grouping together the tests for all of the voter-level sub-groups when adjusting standard errors). Specifically we will test for differential effects along the following dimensions:

- Competitiveness of constituency (primary): the impact of debates on voting choices is expected to be increasing in the competitiveness of the race, as measured by the (decreasing) margin between vote shares for first and second place MP candidates in the previous 2007 election. We will also test (two-sided) whether competitiveness differentially effects measures of policy alignment and likeability.

Deleted: i.e. education and radio ownership are likely positively correlated with general political knowledge

Comment [KC3]: As the analysis of voting returns data involves two additional datasets that we have not yet cleaned, we are separating this part of the plan out (along with corresponding outcomes) and will lodge an updated PAP covering analysis of the electoral data after the exit poll analysis is completed.

Deleted: Additional analysis will use polling-center level voting returns data from the National Electoral Commission (NEC), taking the form: $V_{pc} = \beta_0 + \delta T_{pc} + Z'_{pc}\Gamma + U'_{pc}\Psi + c_p + \varepsilon_{pc}$. . . (2) where V is the outcome (vote share, turnout rate) measured for the polling center p within Parliamentary constituency c ; U is a vector of polling center control variables to be determined from analysis of the control group community survey exit poll data; and other terms remain as above. We will run two specifications: (i) the main specification will include the additional “pure” control polling centers that were not treated nor surveyed in the exit poll and will omit any elements of U that are not available for these centers; and (ii) a robustness check specification that omits the “pure” control centers and includes the full set of polling center characteristics in U .

Comment [KC4]: Our analysis of the I-level data helped us define a more specific measure of competitiveness and candidate performance and narrow the set of outcomes for which we might expect heterogeneous effects.

Deleted: ~~Candidate divergence and competitiveness (primary):~~ the impact of debates should be increasing in the revealed divergence in policy positions and competence of the participating candidates as measured by expert panel and audience rankings from the before/after debate surveys and the interaction of divergence with voting behavior in control polling centers (i.e. if vote shares for the debate winner are already very high in control areas, there is little scope to increase them even if performance in the debate was lopsided). This will involve testing for heterogeneous effects across constituencies (i.e. how the constituency-level ATE varies along key dimensions like ethnic bias) and across two bins of constituencies (i.e. those where one would expect larger versus smaller effects).

- **Candidate performance (primary):** better debate performance (as measured by the expert panel) is expected to enhance the effects of debates on the ability of voters to correctly locate candidate's policy positions, and divergence in performance is expected to enhance treatment effects on votes for the debate winner.
- **Lesser known candidates (secondary):** voter response to strong (weak) performance by less well known candidates (including PMDC, female and non-incumbents) may be stronger than that for other better known candidates, as voters may have greater scope for updating their beliefs
- **Subgroup analysis (primary):** the voting literature suggests that the impact of debates could vary by gender, age, and level of political informedness / naïvete. These tests will be two-sided. We further predict weaker effects for people who do not speak Krio well and may have had trouble understanding the debate.

We further plan to conduct descriptive analysis in the following areas:

- **Spillovers:** establish whether controls saw / heard about the debates in the exit polls; test whether the impact of debates is positive and decreasing in distance from nearest treated polling center.
- **Dissipation of effects:** test whether the impact of debates weakens as the time between the debate screening and Election Day increases
- **Treatment saturation:** test whether the treatment effect is increasing in treatment saturation at the level of polling center; verify that the TOT effect is greater than ITT at the individual level (if some residents of treated polling centers did not attend the screening)
- **Reaction to polling center results:** test whether responses to the exit poll survey systematically vary between those who were surveyed before versus after preliminary results were posted on polling centers
- **Impacts on competitiveness:** calibrate the expected impact of debates on the competitiveness of races if taken to scale, based on estimated TEs on vote shares. Also use voter ratings of Presidential candidates to link and order ratings of MP candidates across constituencies to estimate the impact of sending the best candidate of a given party to other constituencies
- **Priming:** test for survey priming effects by comparing (i) those who attended a screening and were surveyed during the screening (the before and after survey group, plus the after only survey group) to (ii) people who attended a screening but received only an incentive survey (i.e. basic demographics were the only items collected). This complements our analysis in the I-level treatments data that compares surveyed controls to "pure" controls.

Deleted: in voting returns data

Comment [KC5]: Not including priming on the previous draft was just an oversight, as from the beginning, we designed the implementation of surveys to test for these effects

B. Hypotheses and Outcomes

In what follows we organize hypotheses and outcomes into three families: (i) "Vote choice" concerns changes in actual votes cast, which is the ultimate objective yet will be difficult to influence if stronghold candidates that already have significant advantages perform better in the debates; (ii) "Voting knowledge and behavior" reflects the informedness of voting choices and political participation, which are important in their own right and may serve as a necessary but not sufficient step between the status quo and attaining the ultimate objective of changing votes cast; and (iii) "Secondary outcomes" regarding citizen perceptions of politics that are interesting yet less directly linked to the debate experience. Multiple inference corrections will be implemented across outcomes within hypothesis and across hypotheses within family.

- **Vote choice outcomes**

a. Hypothesis 1: Exposure to debates increases **vote shares** for the candidate that performed the best in the debates

i. TE measured by vote choice in exit poll data

ii. Debate winner measured by audience ratings and expert assessment (i.e. two outcomes)

Deleted: (reduces)

Deleted: (worst)

Deleted: (primary test) and in electoral returns (lower power, secondary test)

Deleted: / loser

b. Hypothesis 2: Exposure to debates increases the willingness to **vote across party lines**

i. TE measured by vote choice and ethnicity in exit polls. Two additional measures expand the concept to incorporate non-affiliated tribes: voting for a different party for MP in 2012 than in 2007; and splitting ticket for MP (i.e. party MP different than party Pres)

Deleted: (primary test, limited to members of affiliated tribes); reduced forecasting power of ethnic census shares on electoral returns (lower power, secondary test)

Deleted: (primary)

Deleted: or party LC

- **Voting knowledge and behavior outcomes¹**

a. Hypothesis 3: Exposure to debates increases **political knowledge** and leads to more informed voting

i. TE measured for general political knowledge as mean index on ability to name MP roles, CFF amount, healthcare entitlement, gender equity percentage

ii. TE measured for individual candidate attributes as ability to name candidates, distinguish better educated, public office experience, incumbency and more likely to report personal characteristic as primary determinant of voting choice

iii. TE measured for candidate policy stances by ability to correctly place candidate view on Gender equity, first priority issue, free health care implementation

Deleted: (primary), and mean index on

Deleted: (secondary)

b. Hypothesis 4: Exposure to debates increases **policy alignment**

i. TE measured on the empirical match between voter's position expressed in survey and reported stance of his/her selected candidate expressed in the debate on gender equity, priority issues, and free health care implementation. This alignment can result from persuasion and/or voters choosing aligned candidates, and we aim to unpack the two mechanisms.

ii. We aim to portion out potential drivers of policy alignment along a few dimensions

1. Persuasion vs choosing aligned candidates: MP vote choice in 2007 and Presidential vote choice provide an indication of party loyalty. Greater policy alignment for voters whose 2013 MP choice matches these indicators is more suggestive of persuasion, while those whose choice differs is more suggestive of choosing aligned candidates

2. Learning policy information vs partisan loyalty: Alignment driven purely by an effort to reduce dissonance between party preference and policy stance is less likely for voters from unaffiliated tribes, for less partisan issues (i.e. FHC is much more partisan than GEB or priority issue), and in constituencies where candidates from different parties articulated the

Comment [KC6]: As noted in the I-level PAP, we realized that the outcomes for policy alignment, be it from persuasion or choosing aligned candidates, were empirically equivalent, so have combined the two hypotheses into one and noted how we will attempt to distinguish the two mechanisms.

¹ Note that many of these questions are not relevant for certain candidates and issues in particular constituencies: i) PMDC candidates did not participate in the debates held in constituencies 32, 46 and 47, and did not run for office in 15 and 56; ii) the APC candidate did not participate in the debate in constituency 14; and iii) the free healthcare questions were omitted by the moderator in constituency 85.

same position. By contrast, voters moving into greater alignment with their predicted party position (as predicted by ethnicity or 2007 MP vote or Presidential vote) and the candidates diverged in the debate, is more suggestive of partisan loyalty

- c. Hypothesis 5: Exposure to debates enhances voter openness to other parties
i. TE measured by higher likeability ratings for all candidates (i.e. own party, rival party, and third party where applicable) in exit polls

• Secondary outcomes

- a. Hypothesis 6: Exposure to debates mobilizes the public and leads to greater turnout
i. TE measured by turnout question in exit polls verified by enumerator ID card inspection (primary) and electoral returns (lower power, secondary)
b. Hypothesis 7: Exposure to debates increases the perceived legitimacy of elections
i. TE measured by increasing confidence that elections are free and fair in exit polls
c. Hypothesis 8: Exposure to debates increases interest in politics
i. TE measured by question on frequency of discussing politics in exit poll and ability to name national candidates who were not involved in the debates (presidential candidates, and MP candidates who did not participate in the debates).

↓

3. Domains B and C: Effects of PC-level Debates on Candidate and Party Campaigning

These two domains capture potential effects of the polling center-level screenings on the campaign strategies of candidates and political parties. As candidates and party officials are two different sets of actors we treat them as different domains but combine the exposition of the approach here as it is the same for both.

A. Econometric specification

Analysis of treatment effects will take the form of:

$$Y_{impc} = \beta_0 + \delta T_{pc} + X'_{ipc}\Pi + Z'_{pc}\Gamma + W'_{ipc}\Psi + c_p + \varepsilon_{ipc} \quad (3)$$

where outcome Y (i.e. receiving a gift) is measured for individual i in relation to candidate m where the individual is registered in polling center p within Parliamentary constituency c ; T is an indicator variable equal to one if the polling center received the debate video screening treatment; X is a vector of indicator variables that denote the stratification bin from which exit poll respondents were drawn (where the bins were constructed by age and gender); Z is a vector of indicator variables that denote the stratification bin from which the polling center was drawn (where the bins were constructed by number of registered voters and distance to nearest neighboring center); W is a set of additional control variables that will be determined from analysis of the control group data with an eye toward identifying individual characteristics of political gift receipt; c is a set of constituency-specific fixed effects (the level of debate and candidates); and ε is an idiosyncratic error term clustered at the polling center level. Data concerning candidate expenditure will come from the individual-level exit polls; while exit poll data (and analysis)

Deleted: <#>¶

<#>Hypothesis 4: Exposure to debates mobilizes the public and leads to greater turnout¶

<#>TE measured by turnout question in exit polls (primary) and electoral returns (lower power, secondary)¶

<#>Hypothesis 5: Exposure to debates increases policy alignment¶

<#>TE measured by match between voters position expressed in exit poll question and reported stance of their selected candidate expressed in the debate on gender equity, priority issues, CFF disclosure, and/or free health care implementation. Note that alignment measures will be tailored by constituency to reflect the actual divergence (avoiding a lack thereof) amongst candidates

<#>Hypothesis 6: Exposure to debates persuades voters to adapt their preferred candidate's policy stances¶

<#>TE measured by voter opinion on free healthcare implementation and position on Gender Equality Bill that matches their candidates stance presented in the debate. These measures will be tailored by constituency to reflect candidate positions expressed in the debate, but in most cases we expect APC-(SLPP)-leaning voters to express a more positive (negative) view of FHC implementation compared to their ...

Deleted: 7

Comment [KC7]: As noted in the I-level PAP, we are "demoting" this hypothesis to secondary after seeing the very high turnout rates in the Election (87.3%) as we expect to thus have insufficient power to detect treatment effects. We are retaining it in the analysis as we expect it to remain ...

Comment [KC8]: Cutting the exit poll report of violence as 98.8% of control group responses indicated no violence

Deleted: 8

Deleted: , decreased violence at polling centers (exit poll and NEC incident reports)

Deleted: 9

Comment [KC9]: We are adding this outcome of naming national candidates who did not participate in the debates as an additional test of whether treatment increases political engagement. We will also look at naming local candidates as descriptive analysis to determine whether increase in nation ...

Comment [KC10]: The outcomes under the electoral misconduct hypothesis are in the community level data and NEC official data which have not yet been cleaned, so they will become part of the PAP revision noted in comment 3.

Deleted: <#>Hypothesis 10: Exposure to debates does not increase electoral misconduct¶

<#>Lack of TE documented by questions regarding police presence, inappropriate influence, election officials wearing party ...

for party support will come from (and be conducted at) the community-level. Hypotheses here are two tailed, as candidates and parties could plausibly treat campaign effort/expenditure as a substitute for the screening publicity, or they could compensate for the greater competitiveness of the race by allocating more effort/resources to treatment areas. The main objective of these domains is to test whether campaign investment complements, substitutes or is unresponsive to treatment allocation.

We will further test for heterogeneous effects along three dimensions:

- **Debate performance:** As the response of candidates and parties could vary by how well their candidate performed in the debate, we will test for heterogeneous effects by relative debate performance and degree of performance divergence as measured by expert panel and audience rankings in the before/after debate survey.
- **Party:** budget and strategy may vary by party, so we will test for differential response from the incumbent (APC), opposition (SLPP) and third party (PMDC).
- **Competitiveness:** size of response is likely decreasing in the expected vote margin.

B. Hypotheses and Outcomes

In what follows we organize hypotheses and outcomes by domain (candidate versus party).

- **MP Candidate outcomes**
 - a. Hypothesis 1: Candidate allocation of **campaign effort and expenditure** is responsive to debate publicity
 - i. TE measured by receipt of any campaign gift, type and value of the gift, number of candidate visits in the 6 weeks leading up to the Election
- **Party outcomes**
 - a. Hypothesis 1: Party allocation of **campaign support** is responsive to debate publicity
 - i. TE measured by allocation of number of visits by party officials and party candidates for all races, number of political rallies, number of posters and number of gifts distributed in the community in the 6 weeks leading up to the Election²

Comment [KC11]: Note, this will not be analyzed until the community level data is cleaned, but we are leaving here for clarity of exposition. We will note any revisions coming from what we learn now with respect to candidate spending in the PAP revision referenced in comment 3.

² Note regarding interpretation: there is some overlap between the information collected for the party and the MP above. The questions on the community survey for party cover gifts from party officials and candidates for any office, where the offices are President, MP, Local Councillor and Council Chair.

SIERRA LEONE 2012 ELECTIONS PROJECT

PRE-ANALYSIS PLAN: INDIVIDUAL LEVEL INTERVENTIONS

PIs: Kelly Bidwell (IPA), Katherine Casey (Stanford GSB) and Rachel Glennerster (JPAL MIT)

THIS DRAFT: 15 August 2013

This study examines the impact of providing citizens with information about Parliamentary candidates via structured inter-party debates in the lead up to the Sierra Leone November 2012 Elections. Randomization and treatments were conducted on multiple levels: constituency, polling center and individual (details on sampling and randomization are available in the project's AEA trial registry). This pre-analysis plan governs the analysis of the individual level treatments only. It was written and registered before analysis of the individual treatments data. It incorporates learning from analysis of the before/after screening data within the PC-level treatment sites.

1. Background

Our NGO partner, Search for Common Ground (SFCG), hosted and filmed debates between Parliamentary candidates in 14 constituencies. In order to unpack the distinct types of information revealed by these debates, and the extent to which voters differentially react to such information, we implemented four separate treatment arms and two control arms in a crossed randomized design at the individual level. The arms aim to disentangle four different components of the debate that might impact voting behavior: i) objective information such as candidates' policy platforms and past experience; ii) less tangible indicators of candidate capacity such as candidates' interpersonal and persuasion skills, which signal greater productive ability on the job as an MP; iii) superficial information about candidates such as good looks and wealth; and iv) the social mobilization effect of the community gathering together to discuss politics.

The four treatment arms were as follows: i) the exact same debate screened in the polling-center level treatment but now delivered to an individual; ii) a "getting to know you" video of the same candidates speaking informally about their hobbies and interests; iii) a recording of an independent moderator or journalist summarizing the main policy positions articulated by the candidates during the debates; and iv) a "lab" experiment where respondents make quick inferences after being exposed to pairs of isolated images, short (20 second) video recordings, and a list of candidate names, all collected from MP races outside the respondents' district. The first two treatments were screened to individuals via tablet, the third via audio device and the fourth via a combination of pictures and video recordings on a tablet. At the time of treatment, individuals were surveyed both before and immediately after exposure to the treatment. A fifth control group was also surveyed. Respondents in all five arms were then surveyed again at Election time in the exit polls. We are interested in both short run effects, which are captured in the "After" survey data, as well as longer run effects captured the exit poll data. A sixth control group was not contacted at the time of treatment but was included in the exit poll. This second "pure" control group serves as a robustness check on the concern that the information conveyed by debates does not in fact matter but that instead the surveys and research activity simply increase the salience of the election and encourage greater deliberation in voting choice. We will test for this by comparing the first group of controls, who may be primed to pay attention to the election through the experience of being surveyed, to

these pure controls who are exposed to the same potential spillover effects but who first encounter the research teams in the exit polls after they have voted.

Within the 14 constituencies selected for participation in the debates, we first allocated polling centers to the polling-center level intervention and control group. This PC-level sample drew in VRCs that were sufficiently small (fewer than ~900 registered voters) and far apart from their nearest neighbor (at least ~ one mile). In constituencies where there were a sufficient number of polling centers left over after this process, we allocated the remaining larger and closer together VRCs into the individual-level treatment and control group, stratifying by number of registered voters and distance to nearest neighboring VRC. Within each of the treatment VRCs, households were divided into those with only female registered voters, only male, and both male and female registered voters (based on an earlier household listing exercise). We randomly assigned the six arms to households within each of these bins, and randomly selected respondents within each household to receive the individual-level treatments and/or survey(s).

2. Econometric Specifications

We are interested in the absolute treatment effect of each of the three treatment arms (debate, get to know you and radio report) compared to the control group, as well as the net or relative effect of each treatment arm compared to the other treatments. Analysis of the lab experiment treatment arm will be conducted separately. Analysis will take the following form:

$$Y_{ihtpc} = \beta_0 + \delta_t T_{htpc} + \mathbf{X}'_{hpc} \boldsymbol{\Pi} + \mathbf{Z}'_{pc} \boldsymbol{\Gamma} + \mathbf{W}'_{ihtpc} \boldsymbol{\Psi} + \mathbf{c}_p + \varepsilon_{ihtpc} \quad (1)$$

where outcome Y (i.e. vote choice) is measured for individual i living in household h assigned to treatment arm t registered in polling center p located in Parliamentary constituency c . T is a dummy variable indicating assignment to treatment arm t ; \mathbf{X} is a vector of indicator variables that denote the stratification bin from which the household was drawn (where the bins were determined by the gender composition of registered voters); \mathbf{Z} is a vector of indicator variables that denote the stratification bin from which the polling center was drawn (where the bins were constructed by number of registered voters and distance to nearest neighboring center); \mathbf{W} is a set of additional control variables that will be determined from analysis of the control group data and will vary by hypothesis with an eye toward identifying individual characteristics that do not vary with treatment and that help explain variation in a particular outcome (i.e. education and radio ownership are likely positively correlated with general political knowledge); \mathbf{c} is a set of constituency-specific fixed effects (the level of debate and candidates); and ε is an idiosyncratic error term clustered at the polling center level. Our main specification includes the full set of controls (\mathbf{X} , \mathbf{Z} and \mathbf{W}); we will also show results for the sparser specification that includes only the stratification variables as controls (\mathbf{X} and \mathbf{Z} only) as a robustness check. For each treatment arm, the coefficient of interest is δ_t , the average treatment effect for treatment t compared to the control group. The control group is defined as only the respondents in the first control arm for immediate outcomes in the After survey; and respondents in both the first and “pure” control arms for longer run outcomes in the exit poll survey. Participants in the Lab experiment arm, who were not exposed to any media concerning candidates from their own constituency, will also serve as controls. We will further conduct tests pooling multiple treatment arms together, for example $\delta_{t \in D, R, G} \geq 0$. We will further test a series of hypotheses about the relative effects of the different treatment arms that take the form $\delta_t \geq \delta_{\sim t}$ explained below. For each arm and combination of arms, we will examine immediate effects captured in the After survey as

well as sustainable effects captured in the exit poll survey. Tests will be one-sided in the direction stated in the hypothesis, unless specified as two-sided. We will report treatment effects of all individual outcomes of interest as well as mean effects indices by hypothesis and sub-hypothesis as organized below.

Heterogeneous Effects

We will test for heterogeneous treatment effects at the level of constituency, candidate and voter, adjusting for multiple inference within each level (i.e. grouping together the tests for all of the voter-level sub-groups when adjusting standard errors). Specifically we will test for differential effects along the following dimensions:

- **Competitiveness of constituency:** the impact of debates on voting choices is expected to be increasing in the competitiveness of the race, as measured by the (decreasing) margin between vote shares for first and second place MP candidates in the previous 2007 election. We will also test (two-sided) whether competitiveness differentially effects measures of policy alignment and likeability.
- **Candidate performance:** better debate performance (as measured by the expert panel) is expected to enhance the effects of debates on the ability of voters to correctly locate candidate's policy positions, and divergence in performance is expected to enhance treatment effects on votes for the debate winner.
- **Lesser known candidates (secondary):** voter response to strong (weak) performance by less well known candidates (including PMDC, female and non-incumbents) may be stronger than that for other better known candidates, as voters may have greater scope for updating their beliefs
- **Subgroup analysis:** the voting literature suggests that the impact of debates could vary by gender, age, and level of political informedness / naïvete. These tests will be two-sided. We further predict weaker effects for people who do not speak Krio well and may have had trouble understanding the debate.

Descriptive Analysis

- **Dissipation of effects:** For each arm and hypothesis we will test for dissipation of effects by comparing the size of TE measured in the After survey to its equivalent in the Exit poll.

3. Hypotheses and Outcomes

This part of our research agenda seeks to answer two broad questions: first, whether debates, get to know you and radio reports impact voting behavior; and second, if they do have an impact, what drives the effects of debates - information about policy or persona or both. Part of this second question concerns the speed with which voters can infer information about individual candidates.

Note that the set of outcomes available differs between the after survey and the exit poll, where in general the exit poll is a subset of the After survey (however a few items are only available in the exit poll). The list below is comprehensive of both sets, but tests of short- and longer-run effects will necessarily be limited to outcomes available in the respective datasets.

Notation: Individual level treatment arms are denoted by D = Debate; G = Getting to know you; R = Radio report; and L = Lab.

- **Vote choice outcomes**

- Tests to conduct: $\delta_t \geq 0$ for $t \in D, R, G$; $\delta_{t \in D, R, G} \geq 0$; $\delta_t \neq \delta_{\sim t}$ for $t \in D, R, G$
- a. Hypothesis 1: Exposure to debates increases **vote shares** for the candidate that performed the best in the debates
 - i. TE measured by vote choice
 - ii. Debate winner / loser measured by audience ratings and expert assessment
- b. Hypothesis 2: Exposure to debates increases the willingness to **vote across party lines**
 - i. TE measured by vote choice and ethnicity (limited to members of affiliated tribes)
 - ii. Three additional (also primary) measures expand the concept to incorporate non-affiliated tribes: voting for a different party for MP in 2012 than in 2007; splitting ticket for MP (i.e. party MP different than party Pres or party LC); and voting for a party other than the self-reported party affiliation from before survey.
- c. *Mechanisms of impact*
 - i. ***Policy stance and professional qualifications matter*** for vote choice
 1. As the radio report isolates the impact of policy stances and professional qualifications information conveyed by debates from the persona aspects that are also revealed by debates, $\delta_R \geq 0$ provides evidence in support of the role of these characteristics in voting decisions (Note that an absence of these effects for R does not on its own rule out the value of policy and professional background for vote choice.)
 2. For the debate treatment, we expect to see the impacts of debates on votes for the winning candidate to be increasing in the divergence between candidates in expert assessments of competence (professional qualification and connectedness).
 - ii. ***Candidate persona matters*** for vote choice
 1. As the Get to know you video removes the policy content, $\delta_G \geq 0$ provides evidence in support of the role that candidate persona and likeability play in vote choice

- **Voting knowledge and behavior outcomes**

- a. Tests to conduct: $\delta_t \geq 0$ for $t \in D, R$; $\delta_{t \in D, R} \geq 0$; $\delta_t \neq \delta_{\sim t}$ for $t \in D, R$. Note that G only relevant for subset of outcomes indicated below.
- b. Hypothesis 3: Exposure to debates increases **political knowledge** and leads to more informed voting
 - i. TE measured for general political knowledge includes ability to name MP roles, CFF amount, healthcare entitlement, gender equity percentage, the meaning of MP, who approves the government budget, and exact date of election. Note we do not expect heterogeneous effects on these measures across candidates or constituencies.

- ii. TE measured for individual candidate attributes includes ability to name candidates, distinguish better educated, greater public office experience and older candidates, incumbency and more likely to report personal characteristic as primary determinant of voting choice. Note distinguishing between candidates only applies where the characteristic diverges across candidates in the constituency.
 - 1. Include $t = G$ for naming candidates and distinguishing better educated, more experienced and older candidates.
- iii. TE measured for candidate policy stances by ability to correctly place candidate view on Gender equity, first priority issue, free health care implementation, and CFF transparency; plus local policy issues (where available)
- c. *Mechanism of impact*
 - i. **Comprehension and attention** may vary by mode of information delivery. A finding that $\delta_D > \delta_R$ for general political knowledge questions (H3) suggests that debates may better engage the audience than radio summaries. Check for waning attention by placement of knowledge questions in the program (i.e. MP roles at the beginning, date of election at the end)
 - ii. For D, the impact on correctly locating candidate positions should be increasing in the performance of the candidates in answering policy questions as assessed by the expert panel.
- d. Hypothesis 4: Exposure to debates increases **policy alignment**
 - i. TE measured on the empirical match between voter's position expressed in survey and reported stance of his/her selected candidate expressed in the debate on gender equity, priority issues, and free health care implementation. This alignment can result from persuasion and/or voters choosing aligned candidates, and we aim to unpack the two mechanisms using the before and after data.
- e. *Mechanism of impact:*
 - i. **Persuasion** reflects voters adopting their preferred candidate's policy stances (i.e. because they are convinced by the candidate's arguments, learn something about the policy from the debates, and/or are reducing cognitive dissonance between their candidate and policy preferences). These measures will be tailored by constituency to reflect candidate positions expressed in the debate, but in most cases we expect APC-(SLPP-)leaning voters to express a more positive (negative) view of FHC implementation compared to their counterparts in control areas, and for treated voters to report greater support of the GEB. A before/after test limits the sample to voters whose candidate choice matches their pre-stated party loyalty (i.e. holding party choice constant) and checks for improved alignment in the after survey. Limiting the sample to Independent voters and those of historically unaffiliated tribes is another test, both for policy alignment and more generally for changing voter positions on issues irrespective of vote choice. Power for these tests, however, will likely be low given small sample sizes.
 - ii. **Choosing aligned candidates** reflects moving to a candidate whose policy positions match the voters *ex ante* policy preferences. These measures will be

tailored by constituency to reflect the actual divergence amongst candidates. A before/after test focuses on the voters policy preference stated in the before survey and looks for differential improvements in alignment in constituencies where the candidates diverge on a particular issue compared to constituencies where all candidates agree.

- iii. Finding that $\delta_D > \delta_R$ for measures of policy alignment provides evidence for the impact of ***productive intangibles*** related to MP job performance not easily captured by policy positions and professional background. Finding $\delta_G \geq 0$ for measures of candidate education and professional experience (H3) suggests that these attributes may be conveyed by pattern of speech and presentation.
- f. Hypothesis 5: Exposure to debates enhances **voter openness** to other parties
 - i. TE measured by higher likeability ratings for all candidates (i.e. own party, rival party, and third party where applicable) in exit polls
 - 1. Include $t = G$
 - g. *Mechanism of impact:*
 - i. Finding that $\delta_D > \delta_R$ and $\delta_G > 0$ for measures of candidate **likeability** suggests that voters value the general affability of candidates
- **Thin slice outcomes**
 - a. Hypothesis 10: **Thin slices of information** enable voters to make accurate inferences about individual candidate attributes
 - i. Benchmark test 1: Inferences regarding ***facial competence from photos***: test whether voter inferences about leadership ability and their voting preference after viewing photos predict the winners of MP races with an accuracy greater than chance
 - ii. Benchmark test 2: Inferences regarding ***productive intangibles from public speaking***: test whether voter inferences about leadership ability and their voting preference after viewing video clips predict the winners of debates and MP races with an accuracy greater than chance
 - b. *Mechanism of impact:*
 - i. ***Productive intangibles are more important than purely superficial cues*** for voting preference: accuracy of benchmark test 2 > benchmark test 1
 - ii. If accurate, explore which inferences are most correlated with predictive power, i.e. judge to be better leader vs better looking?
- **Secondary outcomes** See details in PC-level PAP
 - a. Hypothesis 6: Exposure to debates mobilizes the public and leads to greater **turnout**
 - i. TE measured by turnout question in exit polls
 - ii. Note we are “demoting” this hypothesis to secondary since turn out in the actual election was extremely high (87.3%) and we thus expect to have insufficient power to detect treatment effects on this outcome, but are retaining it in the analysis as we expect it to be of general interest
 - b. Hypothesis 7: Exposure to D, G and R increases the perceived **legitimacy** of elections
 - c. Hypothesis 8: Exposure to D, G and R increases **interest in politics**

Additional Secondary Questions

- Hypothesis 12: **Social mobilization**, the community gathering aspect of the polling center-level treatment intensifies the effect of the debate on voter behavior
 - a. TE measured by comparing the TE of the I-level D to the PC-level D. If comprehension and other effects are similar, but other effects are lower for I-level, conclude that the community gathering aspect had a separate effect from the content.
- Hypothesis 13: Effect of treatments is not just **priming** / getting people to focus on the election
 - a. TE measured as no differences between pure controls and surveyed controls in Individual T implementation areas (note: both subject to the same potential spillovers from speaking with neighbors in other T arms, but only the surveyed one is primed)
- Hypothesis 14: Debates do not **increase ethnic voting**
 - a. Check whether exposure to D and G increases the ability of voters to correctly guess individual candidate ethnicity
 - b. In L, check that ability to guess individual ethnicity based on names (which are printed on ballot) and photos (which are posted outside polling stations) is at least as strong as that based on video clips (which are additional information revealed by debates)

SIERRA LEONE 2012 ELECTIONS PROJECT

PRE-ANALYSIS PLAN: ~~POLLING CENTER~~CONSTITUENCY LEVEL INTERVENTIONS

PIs: Kelly Bidwell (JPAL), Katherine Casey (Stanford GSB) and Rachel Glennerster (JPAL)

DATE: 2 June 2014

Updated: 13 March 2015

This study examines the impact of providing citizens with information about Parliamentary candidates via structured inter-party debates in the lead up to the Sierra Leone November 2012 Elections. Earlier PAPs govern the analysis of treatment effects on voter behavior and candidate/party response up through Election Day. This pre-analysis plan sets out the main parameters of analysis for post-Election effects of debates on the behavior of election winners who are now serving in Parliament, over the course of the their first year to 18 months in office. The research design corresponds to the highest level of randomization for this project, where 14 of what we expected to be the 28 most closely contested constituencies were randomly assigned to participation in debates and the remaining 14 were assigned to the control group. This plan was lodged in the AEA trial registry on June 2, 2014, which is before any data analysis began, and before much of the data collection was completed (where the current anticipated completion of fieldwork for the suite of CFF surveys is June 7th, 2014, and data entry from these paper surveys will follow).

A key part of this analysis is an attempt to substantiate on the ground the development expenditures the MPs claimed to have made under their first annual constituency facilitation fund (CFF) and post-election engagement with constituents. The fieldwork plan for this involves triangulation of responses across a number of different respondents and surveys, each of which may be more or less informed and credible, which will complicate data analysis. Our plan moving forward is thus to: i) lodge this initial PAP before field work ends to lock in our main domains of hypothesized effects; ii) enter and clean all survey data; iii) analyze data from the control sample only to generate a more specific list of indicators across surveys and analysis procedures; iv) data enter and code official Parliamentary records by topic; v) lodge an update to this plan with more specific indicators and strategies of triangulation across sources (Note: this is what this update (in track changes mode) does, lodged before any treatment data has been analyzed); and v) conduct analysis of treatment effects. The advance analysis of the control sample (item iii above) aims to accomplish a few objectives. First, we will assess variation in outcome measures and baseline levels of activity, to refine the set of outcomes sensible for analysis. Second, we will test our operating assumptions that residents in the constituency headquarter towns (respondents in the main town CFF surveys) have good information about MP activity throughout the constituency and are relatively similar in their views of MP performance compared to those in smaller villages. To do so, we will use the target village responses as a cross check. If we find that either of these assumptions do not hold empirically, we will devise a strategy for how best to make use of the information collected in the (non-random) target village sample. Third, we will assess how well we can determine the accuracy of responses from different sources. We anticipate that the judgments by our enumerators on the relative truthfulness across respondents will be sufficient here, but if we find that the enumerators were unable to make decisive rankings and that there is considerable differences in view across respondents, then we will develop a strategy for balancing and reconciling conflicting reports.

Key caveat: It is important to note that power at this level is limited due to the small sample, so we will be particularly cautious in interpreting a lack of evidence of treatment effect as suggestive of evidence of no effect in practice. With N=28, we also do not have power to adjust for multiple inference. If we have to pick one hypothesis where we have the strongest *a priori* interest and thus where the per comparison *p*-value is most relevant, it is Hypo D (CFF spending).

1. Background

All candidates in the 28 constituencies were surveyed in the pre-Election period. After that, the control group in the 14 control constituencies were otherwise not contacted by the research team. Candidates in the treatment group were invited to participate in a structured inter-party debate that was moderated and filmed by our NGO partner, Search for Common Ground. The debates were then taken on a polling-center level “road show” in a randomly selected 112 of 224 polling centers plus an additional 85 screenings in satellite villages. We estimate that roughly 19,000 individuals were exposed to this treatment. Additional individual-level screenings were held in a separate set of 40 polling centers. Early in their tenure, winners in the treatment group were shown a video of the debate they participated in, edited down to include only their own statements, to remind them of the commitments they made during the debate and explain how many of their constituents saw the debate via the road show. The research hypothesis is that the publicity of the debates screenings could help solve the candidate commitment problem and thereby enhance the consistency of elected MP behavior with their pre-election promises and generally enhance accountability pressure toward better performance in office.

Second caveat: There are two key ways in which the roadshow treatment is considerably less intense than other obvious ways to disseminate the debates to voters. First, the road show was shown to a relatively small subset of constituents: a back of the envelope calculation would put this figure at around 3%. Broader dissemination, e.g. via radio, would reach many more. Second, the MPs themselves were not present at these screenings, so if the winners did not understand or internalize the number of constituents exposed, it is unlikely to affect their future behavior.

Data for this segment of the analysis draws on multiple sources: i) the official Votes and Proceedings produced by Parliament administration (V&Ps); ii) the official Hansards produced by Parliament administration; iii) committee assignments and minutes of committee meetings, produced by committee clerks; iv) MP candidate pre-election survey; v) winning MP post-election follow-up survey (supplemented with post-survey follow-up phone calls to clarify CFF project locations and expenditures); vi) CFF main community questionnaire; vii) CFF clinic follow-up questionnaire; viii) CFF verification sheet; ix) CFF school follow-up questionnaire; and x) CFF Target village community questionnaire. The first 8 sources apply to all MPs in a standard and equal fashion. The last two sources do not, as they are sampled based on MP reports about the location of school support and general development projects, and are intended primarily as an input into the verification sheet. We will also use information from these latter two sources descriptively to cross check our main assumptions about the level of informedness of main town and clinic respondents and their similarity to more rural constituents.

Econometric specification:

$$Y_{ic} = \beta_0 + \beta_1 T_c + \mathbf{X}_i' \boldsymbol{\Gamma} + \lambda_c + \varepsilon_{ic}$$

where Y_{ic} is outcome for MP i in constituency c , T_c is an indicator signaling that the constituency was assigned to the debates participation treatment, X_i is a vector of MP-level controls {gender, public office experience} chosen for their contribution to increasing R^2 in the control group data, and λ_c are fixed effects for the randomization strata used in the constituency-level assignment (3 bins of raw ethnic bias). The primary specification will include X_i and the robustness check will exclude X_i .

2. Domain D: Effects of PC-level debates on Elected Officials - Hypotheses and data sources

This sections specifies the main areas of hypothesized effects and lists the corresponding sources of data. In general, we will look for effects in both “hard facts,” for example CFF expenditures that are verified via field visits, as well as in MP behavioral or priming responses, where they may be more likely to claim better performance in self-reports.

A. Activity in Parliament

Hypothesis: Accountability pressure of constituent exposure to debates is expected to increase the activity and engagement level of elected MPs. All tests are one-sided towards increased activity. Relevant indicators:

- i) **Attendance** in Parliamentary session as recorded in the V&Ps → specific indicator = percent of all 2012-13 sittings attended
- ii) **Participation** via making public statements in Parliamentary sessions as recorded in the V&Ps and Hansards → specific indicator = total number of comments made over all 2012-13 sittings
- iii) **Committee** membership
 - a. Number of committees serving on as compiled by the Clerk of Parliament → specific indicator = total number of committees as of Q1 2015 report from Clerk of Parliament
 - ~~b. Attendance in committee meetings as recorded by committee clerks. This data will need to be assessed for completeness and accuracy before proceeding with analysis as many MPs serve on multiple committees and recordkeeping may vary substantially across clerks. DROP outcome as attendance records nonexistent for many committees.~~
- iv) **(Secondary) Self-reported** MP activity in follow-up survey regarding discussing topics with other MPs, raising issues during committee meetings, and other promotional work (GEB: Q17, 18, Issue: Q21, 22, 23)

B. Consistency with pre-election promises

Hypothesis: The publicity of the debates helps solve the candidate commitment problem and makes their post-election behavior in Parliament more consistent with their pre-Election promises. All tests are one-sided towards increasing consistency. Relevant indicators:

- i) Participation in **Parliamentary session in key priority areas**, where pre-election priorities were collected in the MP candidate survey and in the debates for treated MPs, and post-election participation is recorded in the V&Ps and Hansards → specific indicator = total number of comments during an agenda item relating to priority sector over all 2012-13 sessions
- ~~ii) Voting in accordance with pre-election stated preferences for the Gender Equity Bill (when it arises in Parliament); for the Freedom of Information Bill; and votes that relate to the sectors~~

- ~~specified as first priority issues as recorded in V&Ps and Hansards (relevant votes need to be identified and coded)~~ DROP outcome due to lack of variation: the GEB has not yet been voted on and the FIB was passed unanimously, as were bills in priority sectors
- ~~iii)ii)~~ **Membership in committees** that govern stated key priority issues → specific indicator = # of committees member of in priority issue areas in Q1 2015 report from Clerk of Parliament
 - ~~iv)iii)~~ **Constituent assessment of consistency** with and **performance in promoting** priority areas in CFF Main Town ~~and Clinic surveys~~ (QC5-C14 in main; ~~QC11-15 in clinic~~) → specific indicator = # respondents saying “yes” that MP focuses on priority sector / total # main town respondents (note: main town assessment of “performance promoting the sector” near perfectly correlates with this measure, so we are including only this one)
 - ~~v)iv)~~ **(Secondary)** Correspondence between **MP self-reports** in pre- and post-Election surveys (GEB Q15, Issue Q19, CFF Q24, Transparency Q25)

C. Constituency engagement

Hypothesis: Accountability pressure of constituent exposure to debates is expected to increase post-election engagement with constituents. All tests are one-sided towards increased engagement. Relevant indicators:

- i) Number of **visits** to constituency as verified across the CFF main community and clinic surveys (QM2-5 in Main and M2-5 in Clinic) → specific indicator = average # of visits over all respondent reports. Code “77 – too many to count” and “resides locally” to the max{5, max mean in full sample}
- ii) Number of **substantive meetings** held with constituents as verified across the CFF main community survey (Main QM6-8, with truthfulness check QM8) → specific indicator = average # meetings over all respondent reports
- ~~iii)iii)~~ **ADD: Overall assessment of doing a good job** → specific indicator 1 = average # of reports of doing a good job in any sector (C7-C14) over all main town respondents; specific indicator 2 = average # of reports of doing a good job in health (C15) over all clinic respondents
- ~~iii)iv)~~ **(Secondary)** Self-reported visits to constituency in the MP follow-up survey (Q28-30)
- ~~iv)v)~~ **(Secondary)** Self-reported number of substantive meetings in the MP follow-up survey (Q31-32)
- ~~v)vi)~~ **(Secondary)** Number of **clinic oversight visits** captured in clinic survey (QC1-5)

D. CFF spending

Hypothesis: Accountability pressure of constituent exposure to debates is expected to increase development expenditure under the CFF (and potentially through mobilizing other funds, TBD). All tests are one-sided towards increased engagement. Relevant indicators:

- i) Overall **proportion of CFF funds spent** that can be verified; and **proportion were dedicated toward constituency development** as opposed to transport. (Note this is one unified indicator as we had no way to independently verify self-reported transport expenditures) Many sources to this, but primary metric should be summarized in the CFF verification sheet. Triangulation from: CFF projects in Main (QC15-17; and Section P); Target (QC18-19, and Section P); contributions to clinic development (QC6-9 in clinic survey); contributions to scholarships and

- school development (QC3-9 in school survey); and note truthfulness assessment questions at end of relevant sections → specific indicator = amount of development funds verified in the field from the verification sheet / 43.8 M leone allocation
- ii) **(Secondary) Self-reported** expenditures in MP follow-up survey (Q33) and post-survey clarification phone calls – proportion reported and proportion for development;
 - iii) **(Secondary) Self-reported biggest accomplishments** as MP in follow-up survey (Q34)

3. Descriptive analysis

These surveys also were designed to collect indicators that flesh out other areas of primarily descriptive analysis.

- A. Descriptive analysis of Target village and School survey responses to questions about MP consistency, MP visits, meetings, and performance
- B. MP self-reports of participation in key areas will be used to select relevant Hansards and cross-check official records in MP follow-up survey: Q13-14 general debate, Q16 GEB, Q20 priority issue, Q26-27 on Freedom of info to both cross check and potentially add nuance to TE estimate of accountability and activity level if find variation in abstention or failed participation attempts
- C. Content or textual analysis of the V&Ps and Hansards

Two other areas relate to earlier stages of the research design but the data for which was collected in the MP follow-up survey that this plan governs.

- D. Secondary data on party response to assignment of treatment and control in MP follow-up survey Q8-9 campaign support
- E. Check on T/C balance and/or (rule out) party selection response to treatment assignment in MP follow-up survey Q10-12 quiz questions. Supplement this with data on candidates in pre-election survey

Appendix Table 1: Domain C - Treatment Effects of Polling Center Screenings on Party Officials

Outcome	Control mean	Treatment effect	Standard error	Naïve p-value (2 sided)	FDR q-value	N
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Hypothesis C1. Mean Effects Index (all 21 outcomes)</i>	<i>0.000</i>	<i>0.082</i>	<i>0.052</i>	<i>0.113</i>		<i>224</i>
Any visits by party officials, APC	0.819	0.004	0.060	0.941	1	210
Number of visits by party officials, APC	1.857	0.427	0.383	0.266	1	210
Any political rallies, APC	0.248	0.045	0.055	0.412	1	207
Number of political rallies, APC	0.467	0.046	0.147	0.753	1	209
Any party officials distributed gifts? APC	0.481	0.080	0.063	0.206	1	205
Number of community members receiving gifts, APC	0.295	0.069	0.044	0.114	1	208
Number of posters displayed in community, APC	0.699	-0.022	0.042	0.601	1	207
Any visits by party officials, PMDC	0.368	0.008	0.069	0.912	1	133
Number of visits by party officials, PMDC	0.515	0.386	0.342	0.260	1	134
Any political rallies, PMDC	0.044	-0.028	0.022	0.198	1	134
Number of political rallies, PMDC	0.044	0.004	0.043	0.934	1	134
Any party officials distributed gifts? PMDC	0.045	-0.003	0.032	0.928	1	133
Number of community members receiving gifts, PMDC	0.025	-0.002	0.017	0.924	1	133
Number of posters displayed in community, PMDC	0.235	0.048	0.052	0.353	1	132
Any visits by party officials, SLPP	0.739	0.051	0.062	0.409	1	222
Number of visits by party officials, SLPP	1.679	0.066	0.230	0.773	1	224
Any political rallies, SLPP	0.159	0.011	0.044	0.809	1	213
Number of political rallies, SLPP	0.315	-0.079	0.104	0.448	1	219
Any party officials distributed gifts? SLPP	0.368	0.078	0.072	0.278	1	213
Number of community members receiving gifts, SLPP	0.226	0.038	0.040	0.339	1	214
Number of posters displayed in community, SLPP	0.555	0.047	0.035	0.187	1	221

Notes: i) significance levels + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$; ii) robust standard errors clustered by polling center; iii) specifications include stratification bins for the polling center (number of registered voters and distance to next nearest) and constituency fixed effects; iv) mean effects index constructed following Kling, Liebman and Katz 2007 and is expressed in standard deviation units; v) adjustments to control false discovery rate (FDR) computed following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008); and vi) data source is the community level exit poll survey.

Appendix Table 2: Treatment Effect Estimates for Secondary Hypotheses

Outcomes and datasets by hypothesis	Control	Treatment	Standard	Naïve	FDR	N
	mean	effect	error	p-value	q-value	
	(1)	(2)	(3)	(4)	(5)	(6)
A6.i. Exposure to debates mobilizes the public and leads to greater turnout (1 outcome)						
Mean effects index, group screening exit polls	0.000	-0.210	0.117	0.963		5,247
Individual turnout verified by voter ID card, group screening exit polls	0.984	-0.026	0.015	0.961	0.490	5,346
Polling center-level turnout, National Electoral Commission official returns	0.791	-0.004	0.008	0.675		224
Mean effects index, individual debate treatment	0.000	0.072	0.036	0.026*		1,595
Mean effects index, individual "get to know you" treatment	0.000	0.110	0.061	0.040*		1,590
Mean effects index, individual "radio report" treatment	0.000	0.024	0.044	0.291		1,592
A6.ii. Exposure to debates increases the perceived legitimacy of elections (1 outcome)						
Mean effects index, group screening exit polls	0.000	-0.092	0.083	0.866		5,247
Individual confidence that election was free and fair, group screening exit polls	0.919	-0.025	0.023	0.858	0.432	5,415
Mean effects index, individual debate treatment	0.000	-0.008	0.050	0.563		1,698
Mean effects index, individual "get to know you" treatment	0.000	0.020	0.042	0.320		1,695
Mean effects index, individual "radio report" treatment	0.000	-0.083	0.055	0.930		1,695
A6.iii. Exposure to debates increases interest in politics (3 outcomes)						
Mean effects index, group screening exit polls	0.000	0.054	0.036	0.071+		5,247
Individual frequency of discussing politics, group screening exit polls	0.243	0.031	0.027	0.124	0.102	5,411
Individual names MP candidate not in the debate, group screening exit polls	0.135	-0.019	0.032	0.718	0.395	1,645
Individual knows Presidential candidates' names (of 2), group screening exit polls	1.360	0.097	0.041	0.009**	0.013*	5,413
Mean effects index, individual debate treatment	0.000	-0.030	0.028	0.853		1,595
Mean effects index, individual "get to know you" treatment	0.000	0.015	0.033	0.332		1,590
Mean effects index, individual "radio report" treatment	0.000	0.037	0.031	0.122		1,592

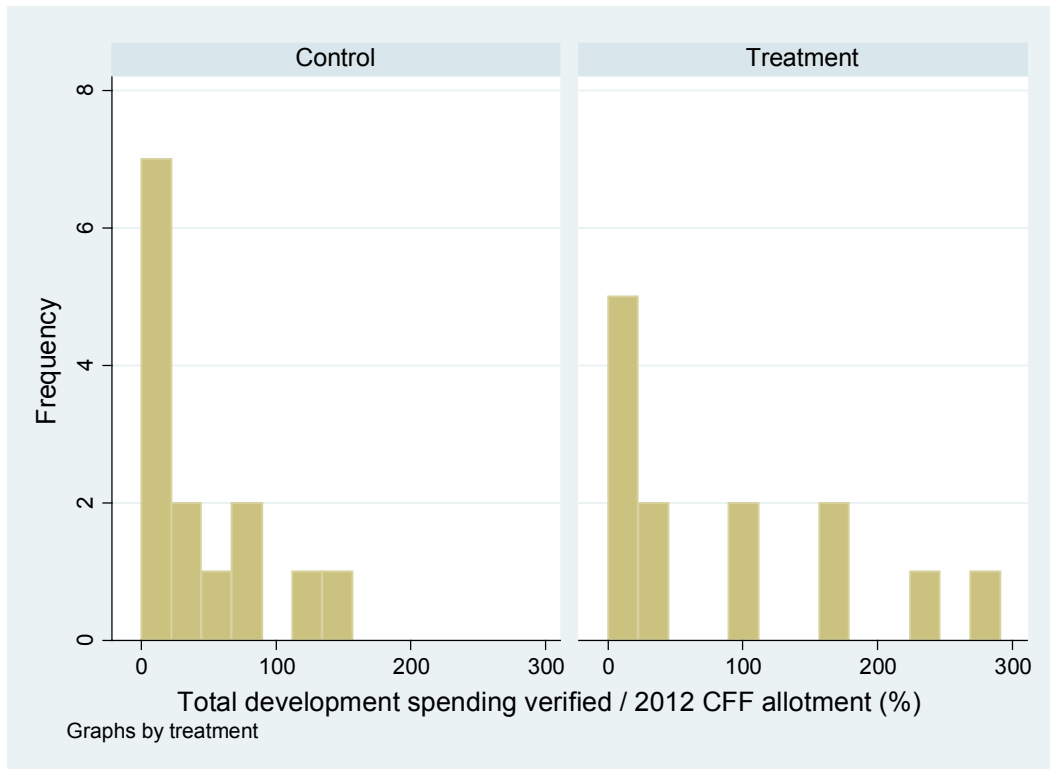
Notes: i) significance levels indicated by + p < 0.10, * p < 0.05, ** p < 0.01; ii) all mean effects indices are expressed in standard deviation units; iii) estimates for mean effects indices in group screening exit polls specified as in Table 1; iv) estimates for individual outcomes in group screening exit polls specified as in Table 2, where false discovery rate (FDR) adjustments apply across all 48 primary and secondary outcomes; and v) estimates for mean effects indices on individual treatments specified as in Table 5.

Appendix Table 3: Treatment Effect Heterogeneity, Subgroup Analysis

Dependent variable = Mean effect index by hypothesis	A1. Political Knowledge	A2. Policy Alignment	A3. Votes for Debate Winner	A4. Votes Across Party Lines	A5. Voter Openness
	(1)	(2)	(3)	(4)	(5)
Female * Treatment	-0.076** (0.021)	0.037 (0.031)	-0.034 (0.039)	0.036 (0.040)	-0.003 (0.030)
Youth * Treatment	0.021 (0.019)	-0.033 (0.031)	0.070 (0.045)	-0.069+ (0.041)	0.011 (0.032)
No Krio * Treatment	-0.023 (0.042)	0.038 (0.061)	0.107+ (0.061)	-0.100+ (0.053)	-0.045 (0.058)
Treatment	0.311** (0.030)	0.095* (0.042)	0.050 (0.050)	0.015 (0.043)	0.096+ (0.050)
Female	-0.062** (0.016)	0.024 (0.027)	0.060+ (0.035)	-0.078* (0.037)	-0.027 (0.029)
Youth	-0.008 (0.014)	0.011 (0.022)	-0.027 (0.033)	0.030 (0.034)	0.002 (0.023)
No Krio	-0.099** (0.025)	-0.022 (0.037)	0.064+ (0.036)	-0.040 (0.040)	-0.004 (0.037)
Observations	5,246				

Notes: i) significance levels indicated by + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$; ii) robust standard errors clustered by polling center; iii) all specifications include stratification bins for the polling center (number of registered voters and distance to next nearest) and constituency fixed effects; iv) specifications further include additional control variables years of schooling, polygamous marital status, farming occupation and radio ownership; v) treatment effects are on the hypothesis-level mean effects index constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units; and vi) data source is the exit poll survey from the group screening sample.

Appendix Figure 1: Histogram of CFF Spending by Treatment Assignment



Notes: This figure plots the distribution of total development expenditures that could be verified on the ground through detailed field visits to each MPs home constituency, scaled by the amount of the 2012 constituency facilitation fund (CFF) allotment, separately for control (on the left) and treated MPs (on the right). Comparing the two subplots shows that the positive treatment effect estimated in Table 5 is driven by differences in both tails: there is a higher frequency of zero values among control MPs, as well as a larger number of high values among treated MPs. Values above one hundred percent reflect the fact that fieldwork occurred after the first 18 months in office (i.e. potentially capturing more than one annual CFF allotment) and that MPs are free to raise additional monies to supplement the CFF.