All Eyes on Them

A Field Experiment on Citizen Oversight and Electoral Integrity*

Natalia Garbiras-Díaz† and Mateo Montenegro‡

May 20, 2021

Abstract

Can ICTs help citizens monitor their elections? We analyze a large-scale field experiment designed to answer this question in Colombia. We leveraged Facebook advertisements to encourage citizen reporting of irregularities and varied whether candidates were informed about the campaign in a subset of municipalities. Total reports, as well as evidence-backed ones, experienced a large increase. Electoral irregularities decreased, and this effect was larger where candidates were informed about the campaign. Finally, the reporting campaign decreased the vote share of candidates dependent on irregularities. We show that this light-touch intervention is substantially more cost-effective than monitoring efforts traditionally used by policymakers.

JEL codes: D72, C93, P16, O17

Key words: Electoral irregularities and fraud, bottom-up monitoring, information and communication technologies, Colombia

*We are grateful for the guidance provided by Daron Acemoglu, Esther Duflo and Ben Olken. This paper has benefited greatly from the conversations with Abhijit Banerjee, Thad Dunning, Leopoldo Fergusson, Ray Fisman, Stuti Khemani, Horacio Larreguy, Monica Matrinez-Bravo, Tara Slough, and all of the participants at the MIT Development and Political Economy lunches, the CPD working group at UC Berkeley and the BUDIE group at CEMFI. We would also like to thank Laura Pulecio, Juliana Barberena and Diana Velazco at the *Procuraduría General de la Nación*, Esteban Salazar at *PARES* and Marlon Pabón, along with the other members of the MOE that helped us out, and without whom this project would have not been possible. Last but not least, we are indebted to César Gutiérrez and Sebastián Cáceres for their amazing help designing the ads used in our interventions, and Estefanía Avedaño for her outstanding research assistance. Funding for this project was generously provided by the J-Pal Governance Initiative, the Center on the Politics of Development (at UC Berkeley) and the George and Obie Schultz Fund. The experiment was approved by MIT's IRB (the *Committee on the Use of Humans as Experimental Subjects*) with reference #1904805455. The RCT is registered in the AEA RCT Registry with unique identifying number "AEARCTR-0004678".

†University of California, Berkeley. Contact: nataliagarbirasdiaz@berkeley.edu.

‡CEMFI. Contact: mateomontenegro@gmail.com, Address: Calle Casado del Alisal, 5, Madrid 28014.

1 Introduction

Clientelism, voter intimidation and electoral fraud are part of the long list of electoral irregularities that persistently threaten democratic institutions in the developing world (World Bank, 2017). Politicians draw on these different strategies, often combining several of them, as a way of distorting elections to their advantage. Beyond the direct consequences of undermining fair elections and eroding political accountability (Hicken, 2011; Stokes, 2005), a growing amount of evidence has shown that different types of electoral irregularities also harm the economic and political well-being of countries.[1]

Despite systematic efforts by governments and the international community to reduce the prevalence of electoral irregularities, these remain pervasive across the developing world. The most widely used and studied measures to deter electoral irregularities, such as deploying electoral observers or auditors, require high state capacity, coordination, and large investments (including, for instance, the training of several hundreds of *in-situ* observers), which are scarce in these contexts, and particularly in remote regions where these issues are more pronounced. At the same time, both the digitization and the information and communication technologies revolution have provided a large number of tools that are cost-effective, scalable, and thus able to overcome those hurdles, which have only recently begun to be exploited in the fight against electoral irregularities.

In this paper, we study whether crowdsourcing the oversight of elections—i.e., outsourcing this task to a large group of people through online platforms—can be effective in the fight against electoral irregularities.[2] We analyze a massive online campaign conducted through Facebook that had the goal to encourage citizen monitoring of elections around the 2019 mayoral elections in Colombia and ask whether it was effective at both generating citizen reporting and reducing electoral irregularities.

This context provides an ideal setting to study this question for three reasons. First, electoral irregularities remain highly pervasive in Colombia despite the presence of domestic and international observers and auditors. Second, in recent years, there have been initiatives by both the Colombian government and NGOs to increase civil society's engagement in overseeing elections through online reporting websites and applications. Finally, internet

[1] By increasing the political returns of targeted transfers, clientelism leads to the under-provision of public goods and it generates public policy inefficiencies (Baland and Robinson, 2007; Khemani, 2015; Vicente and Wantchekon, 2009). Indirectly, it is also correlated to fiscal corruption (Singer, 2009), which in turn might cause inefficiencies for firms and governments alike (Olken and Pande, 2012). Furthermore, voter intimidation might also help to perpetuate violence in weak states (Acemoglu et al., 2013; Robinson and Torvik, 2014).

[2] Although crowdsourcing has been previously used by governments and NGOs to monitor elections, there have not been experimental studies testing for their effectiveness. A prominent example is the case of *Uchaguzi*, an election monitoring digital platform used in Kenya, which has been credited with reducing electoral violence (World Bank, 2016).

access as well as social media usage have been on the rise throughout the country—with 69% of the population connected and over 60% of them registered as Facebook users.

Leveraging the features of this setting, we launched a massive Facebook advertisement campaign—reaching more than 4.4 million citizens, which represents 31% of the targeted population. We promoted the use of an existing online website where citizens can report electoral irregularities, hosted by the *Misión de Observación Electoral* (MOE), a local well-reputed NGO. We randomized a subset of our sample of 698 municipalities (more than half in the country) to receive advertisements encouraging citizen reports.

In designing the advertisements we sent to citizens, we tried to target some of the underlying reasons why citizens might shy away from reporting—which could be critical in the campaign's success. As emphasized by the literature studying persuasion (DellaVigna and Gentzkow, 2010), media campaigns have a direct emotional appeal over viewers, which in our setting might prime intrinsic motivations to spur citizens to report. Alternatively, despite being encouraged to report, citizens may face a "last-mile problem" due to the difficulty of reporting and acquiring the information to do so. To understand the role of these different channels, we designed three versions of the advertisements that contained either: i) a message informing citizens about the MOE's reporting website and providing them with a link to their online form, ii) a call-to-action message urging citizens to act against irregularities by reporting them (but without any information about how and where to do so), or iii) a combination of both.

In addition to understanding citizens' responses to the reporting campaign, we also wanted to study how candidates react to it. To do so, we further cross-randomized whether candidates running for mayor in certain municipalities, as well as their staffers, received a letter informing them about the reporting campaign or not. This second treatment arm allows us to study the general equilibrium effect of these interventions to the extent that, aside from citizens, candidates also become aware of the reporting campaign and have time to react to it. In particular, we expect that this information would deter them from engaging in electoral irregularities, given that the campaign might increase the monitoring of their actions. Ours is one of the few studies that experimentally vary candidates' and citizens' involvement separately when examining interventions around elections.

We find that citizens in the municipalities exposed to the reporting campaign were significantly more likely to file reports, and when they did so, they were also more likely to include information useful for prosecuting the instigators of electoral irregularities. In particular, being exposed to the reporting campaign increases the proportion of municipalities with at least one report by about 9 percentage points, and the proportion of them with evidence-backed reports by 7 percentage points (corresponding to an increase of 33% and

44%, respectively, from the control mean). We further find significant differences in exposing citizens to ads providing information about the reporting website compared to the ads only containing messages encouraging citizens to report, with the former having substantially stronger effects on reporting. This suggests that an important barrier citizens face to reporting in this context is a lack of information about where to do so.

This first set of results suggests citizen monitoring of elections, in fact, increased in the presence of the reporting campaign. But did this discourage candidates from engaging in electoral malpractice? To answer this second question, we constructed two mutually complementary proxies to measure electoral irregularities. First, we created an original database based on media coverage of electoral irregularities at the municipal level from local and national media outlets in a wide array of formats—including written, TV and radio news, both online and offline. To ensure objectivity, we relied on a third-party firm specialized in news analysis and independent coding. Second, drawing from the literature on electoral forensics, we constructed a measure of electoral irregularities based on administrative voting records. In particular, we leverage fine-grain voting record data to proxy electoral manipulation as deviations in the distribution of the tallies' second-digit obtained by each candidate from Benford's 2nd Digit Law, a popular measure in this literature (Hicken and Mebane, 2017).

Results using both of these measures suggest that the increase in citizen monitoring significantly deterred irregularities, even among municipalities where mayoral candidates did not receive the letter: municipalities exposed to the reporting campaign experienced a reduction of 8 percentage points in the likelihood of having an irregularity mentioned in the media, and it decreased the probability that the voting counts deviate from Benford's 2nd Digit Law by approximately 11 percentage points. In examining the types of irregularities affected, we find that the largest decline comes from instances of vote-buying reported in the media—which is one of the most common types of irregularities in this context. Moreover, we also find evidence that the municipalities where we sent a letter to candidates informing them about the campaign reported an extra decline in irregularities—suggesting that candidates' reaction to the campaign is a major factor explaining the observed drop in irregularities due to the campaign.

Next, we study whether this reduction in electoral irregularities might have affected the electoral prospects of candidates whose votes depended on electoral malpractice. To identify such candidates from the large set of those running in the mayoral elections, we conducted a pre-treatment survey three weeks before the intervention. For every candidate running in their municipality, we asked citizens a set of questions related to each candidate's perceived likelihood of engaging in electoral irregularities. We then validated this variable against objective measures of candidates' engagement in illegal activities. The results indicate that the reporting campaign reduced the vote share of these candidates substantially. For instance, we find that candidates who rated above average in the survey-based measure of electoral irregularities experienced a 3.8 percentage point decrease in their vote share in municipalities exposed to the reporting campaign. We further provide evidence showing that this effect is not driven by a difference in the popularity of candidates identified as more prone to irregularities.

The detected drop in electoral irregularities due to the reporting campaign could explain the reduction in the vote share of those candidates who would have likely engaged in such irregularities. However, a potential alternative channel might have been a change in citizens' preferences for different types of candidates. To approximate the relative importance of the first mechanism, which we can identify with our experimental design, we perform a set of bounding exercises that indicate that at least 11%, and potentially 66% of the effect on candidates' vote share is due to the drop in irregularities.

Finally, we show that the reporting campaign was very cost-effective compared to other interventions studied in the literature. A simple back-of-the envelope calculation suggests that the intervention reduced the vote share of the candidates most prone to engage in electoral irregularities at a cost of $0.45 USD for a one percentage point change in their vote share per polling station. In comparison, the more traditional strategy of deploying electoral observers costs several orders of magnitude more (over $500 USD also for a one percentage point change in vote share per polling station, in our most "optimistic" calculations), and this is true even for more recent innovations, which use ICTs to enhance electoral monitoring (Callen et al., 2016; Callen and Long, 2015).

This paper makes several contributions and builds on at least three strands of literature. First, it speaks to extant work on ways to fight electoral irregularities. Within this literature, most studies have focused on either top-down methods to monitor elections or on bottom-up approaches to educate civil society about the negative consequences of electoral malpractice. With regards to the first approach, a longstanding tradition has studied the effect of domestic and international observers, finding mixed results about their consequences.[3] A few recent papers have examined the use of technological innovations to monitor voting aggregates (Callen et al., 2016; Callen and Long, 2015). In terms of bottom-up initiatives, scholars have investigated whether informing citizens about the societal costs and the illegal character of clientelism and other irregularities can bring them to sanction candidates and parties engaged

[3]While many papers document electoral observers reduce irregularities (Enikolopov et al., 2013; Hyde, 2007, 2010; Leeffers and Vicente, 2019), some others point to mitigated effects due to spillovers (Asunka et al., 2019; Ichino and Schündeln, 2012), and yet, a third group has emphasized their negative consequences (Kelley, 2012; Simpser and Donno, 2019).

in those types of misbehavior.[4] We contribute to this literature by studying the effectiveness of a citizen monitoring campaign, which constitutes a third approach that combines elements from both the mobilization and the monitoring strategies and that has not been properly studied to our knowledge.[5] Furthermore, we show that this approach is more cost-effective than the previously-studied alternatives, and we highlight its potential to be fully scalable.

Second, we also contribute to the growing literature examining the use of bottom-up technologies to improve the accountability of governments and the delivery of public goods and services. Spurred by the World Bank's 2004 *World Development Report*, governments and NGOs alike have heeded the call to use these types of technologies in areas as diverse as education, health, public works and elections. Most of the papers in this strand of the literature[6] have focused on studying offline interventions that do not exploit the potential of ICTs and social media to create engagement and to facilitate the spread of information to increase citizen oversight effectively. Despite recent calls to study this "second generation" of bottom-up technologies (Peixoto and Fox, 2016), our paper is the first one, to our knowledge, to do so using field experiments.

Related to this last point, this paper also speaks to the literature on e-governance and ICT interventions designed to increase the accountability and the efficiency of public services. While most of this literature has focused on studying top-down efforts to increase data availability or decrease leakages,[7] this paper focuses on crowdsourcing, which is an understudied technology. Importantly, our findings provide a rare example of how social media can be used to enhance democracy and transparency – in contrast to the preponderant view that has focused on its negative effects (for a review, see Zhuravskaya et al. 2020).

The remainder of this paper is organized as follows. Section 2 provides background information about the intervention's context, including a discussion of the most common types of electoral irregularities in Colombia, an overview of online electoral reporting, and the elections around which the experiment was conducted. Next, Section 3 provides a description of the experimental design, the study sample and the data used. Section 4 presents and

[4]These include campaigns against vote-buying (Blattman et al., 2019; Hicken et al., 2018; Vasudevan, 2019; Vicente, 2014) or electoral violence (Collier and Vicente, 2013)

[5]A few papers study interventions which included electoral reporting components, but they do so only tangentially and their focus does align with our paper's approach. In particular, Driscoll and Hidalgo (2014) show that an information campaign aimed at educating citizens about how to file formal complaints about electoral irregularities around elections in Georgia increased electoral irregularity reports but depressed turnout. They interpret these results as a consequence of citizens' belief that they were being monitored by either the regime or researchers and that retaliation might ensue. Similarly, Gonzalez (2021) finds that areas in Afghanistan with access to cellphone coverage present less electoral fraud and argues that this is due to greater use of an electoral irregularity reporting hotline.

[6]See Fox (2015) for a good review of this literature.

[7]See Banerjee et al. (2020), Callen et al. (2020), Lewis-Faupel et al. (2016), and Muralidharan et al. (2016) for recent examples.

discusses the main results, and then Section 5 provides a cost-benefit analysis of the intervention. Finally, Section 6 considers the relevance of the findings from a policy perspective and concludes.

2 Context

2.1 Electoral Irregularities in Colombia

Electoral irregularities take many forms and permeate every election in Colombia's democracy. Most existing studies have explored their extent and their main features drawing mainly on qualitative accounts.[8] However, only a few studies have documented the extent of the problem quantitatively. Fergusson et al. (2017) use a list experiment to estimate Colombians' engagement in clientelistic practices—broadly defined as receiving particularistic benefits in exchange for their vote or political support. They find that approximately 18% do so at some point in their lives,[9] and this number is larger for rural and poor respondents. Using this same method, Garcia and Pantoja (2015) show that about 7% of voters were intimidated to vote in a particular way in the 2014 presidential elections.

We will be using the term "electoral irregularities" to refer to any conduct affecting elections penalized by Colombian law. Apart from vote-buying and voter intimidation, common irregularities[10] include:

Illicit political advertising: Political advertisement is forbidden on election day and is only allowed within the three months prior to election day. It is also forbidden to place ads on public infrastructure such as light posts or monuments.

Campaigning by public servants: It occurs when public servants use their position to interfere in elections by either trying to favor or harm a particular candidate or party.

Fraud in voter registration: It occurs when citizens register to vote in a polling station located in a municipality or district different from their place of residence in order to obtain an illicit profit or to alter electoral results. This is usually done as a way to facilitate vote buying as explained below.

Election disruption: It happens when someone disturbs or prevents the normal devel-

[8]See for instance, Holland and Palmer-Rubin (2015) and Leal Buitrago and Dávila-Ladrón de Guevara (1990).

[9]Other studies conducted in other countries in Latin America, such as Gonzalez-Ocantos et al. (2012), show similar estimates.

[10]We approximate how common these irregularities are by the number of reports gathered about them. Figure A1 displays the number of reports made to the MOE in the 2015 elections, using the same window of time around election day used for our intervention in the 2019 elections.

opment of elections or vote counting by deceit or force. In practice it frequently occurs through riots led by citizens or candidates.

Electoral fraud: It occurs when electoral results are altered in other ways, such as ballot stuffing.

The organizational details about how electoral irregularities occur vary according to the type of irregularity. As has been reported and studied in diverse contexts (Stokes, 2005; Stokes et al., 2013), vote-buying and other forms of clientelism are carried out in Colombia via local brokers that intermediate between political organizations and voters. These brokers play the important role of providing political organizations with the information required to target and recruit potential voters into clientelistic relationships, as well as ensuring that these voters actually vote in the intended way. A common form of monitoring compliance is by registering voters in polling stations outside of their place of residence so that brokers can control their clients' vote choice (e.g., Rueda 2017).[11]

Other types of electoral irregularities, such as illicit political advertising and campaigning by public servants, are commonly done by or with the complicity of local politicians, who collude with running candidates to return political favors (Arenas, 2018). Voter intimidation is commonly performed by armed actors such as guerrillas, paramilitaries, criminal gangs, or even the military, in collusion with local or national politicians (Acemoglu et al., 2013), but also by non-armed actors, such as employers who threaten their employees to lose their jobs if they fail to vote in a particular way.

2.2 Electoral Oversight and Reporting in Colombia

Several governmental agencies and NGOs run online electoral reporting websites in Colombia. The *Misión de Observación Electoral* (MOE), an independent, non-partisan NGO that is also one of the largest devoted to promoting electoral integrity,[12] hosts one of the most important of these websites, called *Pilas con el voto* (translated roughly as "keep an eye on your vote"). This website allows users to submit reports anonymously, and it only requires users to specify the date and municipality of the reported irregularity. Additionally, it asks the users to describe the reported events in a free-form field, which is only afterward classified by MOE staff as a report about one (or several) of the electoral irregularities and they rank their 'trustworthiness' in three categories (high, medium, and low). This classification depends on

[11] Evidence from other contexts shows that reciprocity (Finan and Schechter, 2012) or, alternatively, targeting voters likely to sympathize with the candidate supported by the broker (Nichter, 2008) are also prevalent methods to ensure compliance in vote-buying.

[12] The MOE has offices in more than half of the municipalities in the country, spread across all 32 departments.

the amount of evidence and facts (places, names, and proof such as videos) provided about the electoral irregularities reported.

As an NGO, the MOE does not have the power to directly investigate and take legal action about the reports they receive. Instead, it prepares official reports based on the information provided by citizens, and redirects them to the government agencies in charge of investigating and sanctioning electoral irregularities. The MOE's reporting website has been so popular that it has been responsible for more than 80% of the reports collected by the central government in recent years.[13]

One of the main agencies involved in both investigating and sanctioning the electoral misbehavior reported to MOE is the Attorney-Inspector General's Office in Colombia (henceforth, AG).[14] The AG is an independent institution that oversees the correct conduct of public servants through both preventive and sanctioning faculties. During the electoral process, it ensures candidates abide by the law and are in good standing to run for office. Additionally, it closely monitors the electoral process.

How effective are watchdog intuitions in investigating and prosecuting the irregularities reported? In this context, answering this question is complicated because a significant fraction of reports do not contain enough evidence for the electoral watchdogs to start a judicial case. For instance, only 13% of reports submitted to MOE were deemed to be of a high level of 'trustworthiness' (meaning they contain hard evidence about the facts reported). Notwithstanding this difficulty, a brief glance at the AG's database of historical reports in the 2010-2018 period suggests two important facts: (1) 2.5% of reports ultimately reached a judicial decision, such a sanction or acquittal; and (2) an additional 12.4% of reports were being investigated by the AG at the end of the period. Since investigations are publicly conducted, and thus might harm the reputation of candidates and parties, we consider these statistics to reflect that reporting does entail substantial costs for candidates accused of committing irregularities.

Finally, it is important to mention that in recent years the MOE has promoted the use of its reporting website through advertisement campaigns conducted by both traditional outlets such as the radio, TV, or movie theaters, but also through social media. One of the rationales for doing so is because they believe that one of the main obstacles for citizens to report has to do with the fact that they ignore how and where they can do so. In line with this hypothesis, data gathered by the MOE shows that one of the most common reasons

[13]Reports redirected by the MOE represented 83% of reports held by the government's unified reporting unit (URIEL) in 2015 and 90.6% in 2011.

[14]The Spanish name for this institution is the *Procuraduría General de la Nación*. It should be distinguished from a second institution called the *Fiscalía General de la Nación*, which is also commonly translated as the Attorney General.

preventing citizens from reporting electoral irregularities to competent agencies is that they lack information about where to do so (Misión de Observación Electoral, 2018).

2.3 The 2019 Mayoral Elections

In Colombia, local authorities—such as mayors, governors, and council members—are elected every four years in a single round under a simple majority rule. For this study, we focus on the 2019 local elections and, particularly, on mayoral races. We decided to focus on mayors since this post is assigned at the municipal level, corresponding to the level at which we randomized our intervention—compared to governors who are elected at the department level. Additionally, the number of mayoral candidates is tractable compared to the council members, which often had several dozen candidates.

In this context, reelection is not allowed, which implies that there are no incumbents in any of the races. Furthermore, given the country's weakly institutionalized party system, and despite the existing electoral rules that seek to reduce party-switching, it is not straightforward to map candidates to parties to make a case for party-wise reelection. This lack of party discipline, especially at the local level, is reflected in the multiplicity of candidates (for instance, the average number of candidates is five in our sample of municipalities).

3 Research Design

This section describes the treatment arms involved in our experiment and the sample of municipalities included. Section 3.1 provides a description of the experiment, Section 3.2 describes the study sample, Section 3.3 reports how randomization was carried out along with the corresponding balance checks, and Section 3.4 describes the scale of the advertisement campaign and provides some measures of users' engagement with the ads.

3.1 Experimental Design

The main intervention consisted of a large-scale Facebook advertisement campaign designed to encourage citizen reporting. The campaign lasted for five days and targeted all Facebook users 18 years or older. It started on October 24, three days before the elections, and it ended on the night of October 28, one day after the elections. This extra day potentially allowed citizens who had witnessed electoral irregularities but had not reported them on election day to report them the next day through the reporting website.

We designed four versions of ads to flesh out the underlying mechanisms through which the campaign could potentially affect reporting. Namely, we randomized municipalities into the following treatment conditions:

C. Placebo Control Group: Municipalities in this group received a Facebook ad containing a 'placebo' message, reminding people about the coming elections – *"Don't forget that local elections will take place on Sunday, October 27"*.

TI. Information message: Municipalities in this group received a Facebook ad informing them about the existence of the MOE's reporting website – *"The MOE has the following website where you can report electoral irregularities: [LINK]. Don't forget that local elections will take place on Sunday, October 27"*.

TC. call-to-action message: Municipalities in this group received a Facebook ad with a call-to-action to report electoral corruption and act against it – *"In these elections let's stop electoral irregularities. Report them! Don't forget that local elections will take place on Sunday, October 27"*.

TI+C. Information + call-to-action message: Municipalities in this group received a Facebook ad containing both messages in TI and TC.

The rationale for each of these experimental groups is the following. We include the placebo message in the control group to net out the effect of politically-oriented advertisement on citizens' behavior. Treatments TI, TC, and TI+C sought to separate two mechanisms through which the ads could have impacted citizens' propensity to monitor elections: that is, by affecting either (1) the cost of reporting, reduced by informing citizens about the reporting website; or (2) by highlighting the urgency to take action against electoral irregularities by reporting them.

In each of these treatment groups, the corresponding text is featured as the header of the advertisement. Additionally, a short video in a slide-show format was shown beneath the header highlighting the most important points of each message. Figure 3 depicts the slides used for the different ads. As depicted, the main background was white so that it did not reflect any of the colors associated with particular candidates, and the main image shows a ballot box with a text urging to vote, which was the main message transmitted through the placebo message.

By targeting ads to the universe of (adult) Facebook users in a municipality, both citizens and members of the candidates' campaigns might have become aware of the monitoring campaign directly. Moreover, they might have also learned about the campaign indirectly, by hearing about it through others. As such, the effects of treatments TI, TC, and TI+C combine both citizens' and candidates' responses to the campaign. However, it is important to notice that candidates' responses were limited by the fact that we sent the ads only three days before the elections.

To understand more fully how candidates would react to this campaign with more foresight, we further cross-randomized whether candidates were informed about the intervention two months before its occurrence through letters. In particular, this intervention allows us to understand if candidates might respond to knowledge about the campaign by reducing their engagement in electoral irregularities given the increased monitoring of elections.

Previous literature has highlighted, however, that knowledge about reporting mechanisms can also be used by monitored agents to engage in *signal jamming*—i.e. in this case, this could occur if candidates indiscriminately reported one another to make the reporting website useless as a way of identifying offenders. In order to tease out this potential issue, in our second treatment arm, we designed three different conditions that we assigned to municipalities receiving any ad (i.e., treatment conditions TI, TC, and TI+C):[15]

TLP. Letter to Candidates with Partial Knowledge: All of the candidates running for Mayor in the municipalities in this group were informed about the monitoring campaign, but they were not informed about the specific reporting website advertised in the campaign.

TLF. Letter to Candidates with Full Knowledge: All of the candidates running for Mayor in the municipalities in this group were informed about the monitoring campaign, as well as the specific reporting website advertised in the campaign.

TC. No Letter: None of the candidates running for Mayor in the municipalities in this group were be informed about the monitoring campaign.

By leaving out information about the precise reporting channel used in TLF compared to TLP, we can isolate the effects of signal jamming to those of simply being informed about the campaign with foresight.

We partnered with the Attorney General's Office (AG) in sending these letters. The reasons for doing so were twofold. First, in Colombia, candidates' and their staffs' contact information is not public. Therefore, the AG helped by providing this information. Second, sending the letters on behalf of the AG maximized the chances that candidates would actually read them. For reference, Figure A2 shows the actual letter we sent to candidates in TLP.

To maximize the effectiveness of these interventions, we sent the letter to candidates approximately two months before the elections and then sent a reminder three weeks before.[16]

[15] Notice that we did not allow the ad control group to be included in this second treatment arm. We made this decision to avoid deception among that group— i.e., informing them about a campaign that was not taking place in that municipality.

[16] As discussed in Section 2.1, the planning and execution of electoral irregularities can take place several weeks and even months before the election, so this additional forewarning time about the campaign would potentially have an important deterrence effect over candidates.

We sent both physical letters and emails to maximize the chances of getting the candidates' attention. Importantly, the deadline to register for candidacy occurred more than a month before we sent the first of these letters, so that this treatment did not generate any differential effect on the selection of candidates.

Figure 1 summarizes the full factorial design of our experiment, and Figure 2 shows the timeline of the different interventions and electoral milestones.

We intentionally designed both interventions not to affect any candidate or party in specific. On the one hand, all ads were non-partisan, and, on the other, we sent the letter to the universe of the mayoral candidates running in the treated municipalities. Thus, we do not have *a priori* reasons to believe that we intervened in the political outcomes of interest in ways different than potentially dampening the electoral prospects of those politicians likely to engage in electoral irregularities.

3.2 Study Sample

The study sample consisted of 698 municipalities coming from every Colombian department (see Figure A3), which contain approximately 19 million inhabitants—almost 40% of the population in the country.[17] These municipalities were chosen from the universe of over $1,100$ municipalities in Colombia following population-based criteria. Namely, we only included the subset of municipalities with a population of at least $5,000$ and no more than $97,000$ inhabitants. These population cutoffs followed two guiding principles. First, we used a lower population bound because Facebook's API does not allow the targeting populations with too few users. Second, we chose the upper bound to keep the costs of the ads within our budget, as we wanted to ensure a high-impact treatment able to reach at least 30% of the population in each municipality based on estimations using data from pilots.[18]

Table A1 presents the summary statistics for a selected set of variables for the municipalities included in the study sample and compares them to the universe of municipalities in the country. As expected, given the selection criteria for our sample, the average population is smaller in our sample ($\approx 27,000$) than in the whole country ($\approx 37,000$). The municipalities in our sample also have a relatively lower Facebook penetration rate, with an average of approximately 41% of the population reported as active users by Facebook, while the average penetration rate across the country is 61%. However, apart from these differences, the experimental sample is fairly similar to the country's average across other characteristics,

[17] There are two main reasons why we choose to use municipalities as the units of randomization. First, because it is finest geographical level at which MOE have systematic data on reports. Second, we opted to treat the universe of mayoral candidates running in a municipality to minimize the risk of favoring a particular party or candidate.

[18] Additionally, we choose not to consider a few municipalities where we had run a pilot in 2018 and where ads had not reached more than 5% of the users as reported by Facebook.

such as GDP per capita, rurality rates, poverty rates and previous reporting behaviour.

3.3 Randomization in Practice: Stratification and Balance Checks

To increase the balance on potential confounders across treatment conditions, we conducted a stratified randomization. We defined strata by the intersection of bins partitioning the sample in three ways: (1) by the 50% and 85% percentiles of the population over the age of 18; (2) by the 20% and 80% percentiles of voter turnout in the first round of presidential elections in 2018; and (3) by whether the municipalities filed reports through the MOE's website around the congressional elections of 2018 above or below the median.

Table A2 reports balance checks for the different treatment arms using five different sets of covariates, while Figure 4 reports the p-values for a subset of comparisons across treatment arms for succinctness. The first set of covariates includes measures of past reporting through the MOE's website. The second set includes socioeconomic covariates, such as population or per capita income. The third set includes political covariates, such as past turnout, but also past voting behavior for different parties. The fourth set consists of region dummies. Finally, we also include covariates of interest such as the share of the adult population reached by the ads and the responses to a pre-treatment survey we discuss in the next sections.

The results suggest that municipalities are well balanced across treatment arms. Only seven differences in means out of 248 comparisons in Table A2 are statistically significant at a 10% level or less. While these imbalances might have only arisen by chance, this justifies including covariates in our main specifications as a robustness check (in Section 3.6 we explain how we incorporate them in detail).

3.4 Ad Campaign Scale and Engagement

Table 1 provides summary statistics of several measures reflecting the scale of the ad campaign i) aggregated across all of the municipalities in the sample, ii) per municipality, and iii) per population of age 18 years and above (i.e., those eligible to vote).[19]

Overall, the ad campaign was successful in reaching a large population in the targeted municipalities. Approximately 4.4 million Facebook users saw the ad corresponding to their treatment condition an average of ≈ 3 times. In each municipality, the ad reached, on average, $6,245$ people, which represents 23% of the population in these municipalities and 31% of the registered voters in that election. Additionally, our ads were successful in generating substantial engagement by users: more than $23,000$ people clicked on the link to the MOE's reporting website—i.e., an average of 33.5 people per municipality—at a cost of

[19]We took the information contained in this table from Facebook's Marketing API. For ease of interpretation, we altered the original name of some of the variables. In particular, 'viewers of the ad' and 'times the ad appeared on a screen' corresponds to Facebook's 'unique reach' and 'impressions' variables.

approximately $0.23 per link click. In terms of other types of engagement, in the average municipality, 14 people 'reacted' to the ads (i.e. by 'liking' it, 'loving' it, etc.), 6.5 of them shared them, and 0.63 of them left a comment.

While these statistics indicate that the ad campaign was successful in producing engagement and reaching a wide audience across the *average* municipality in our sample, it did not do so across individual municipalities. For instance, in the extreme, there were 21 municipalities where absolutely no user viewed them. The reasons for these differences in the viewers of the ad might have included many external factors, such as problems with internet connectivity in different areas, or errors in the geolocation of the municipality by Facebook. However, these differences do not represent a threat to the identification strategy given that they are well balanced across treatment arms: Panel E of Table A2 shows that there is balance in terms of municipalities with no viewers, in the number of viewers and also in the percentage of the population who viewed the ads. In Section 3.6, we come back to the issue when we discuss our estimation strategy for the intervention's effects.

3.5 Data

3.5.1 Outcome Variables

Our main outcomes of interest are the reports submitted by citizens to the MOE's reporting website and (proxies) for the actual occurrence of electoral irregularities. Additionally, we also analyze the impact that the interventions might have had on electoral outcomes. We measure these outcomes using a mix of administrative and originally collected data. Below, we describe in detail how we measure each of these outcomes.

1. Reporting. Our main outcome variable to assess whether this campaign was successful in getting citizens to report is the number of reports per municipality collected through the MOE's website. We only consider reports submitted on the dates of the intervention: i.e., from October 24, when the campaign started, to October 28, when it ended. In addition to analyzing the total number of reports, we also disaggregate them by their 'trustworthiness,' assessed and classified by the MOE as either high, medium, or low, depending on the evidence and the information contained in the reports about the electoral irregularities, as explained in Section 2. We define reports as "high quality" if these are classified as either of a medium or high level of trustworthiness.[20] As such, we can test whether the campaign successfully manages to induce useful, evidence-backed reports that ultimately can put checks on misbehavior by politicians or if it only affects the margin of low-quality reports.

[20] We do not distinguish medium and high trustworthiness reports, since there are very few reports of the latter type in our sample (under 10% of municipalities in the sample submit such reports).

2. Electoral irregularities. We use two different types of measures to proxy for the actual occurrence of electoral irregularities.

First, we capture irregularities by their mention in the news. There are two main challenges in using this type of measure to proxy for actual electoral irregularities. First, municipalities included in our sample are small and so do not attract as much attention from major news outlets as bigger ones. Second, some news might originate from citizens' reports through the MOE, which could bias our estimates in the direction of finding more news about irregularities in places receiving the treatment.

To overcome these difficulties, we construct an original database of electoral irregularities covering both local and national media, and from a large set of media types, such as TV, radio, and both print and online newspapers. This focus on several types of news allows us to maximize the chances of including news about the municipalities in our sample. Furthermore, we purge from our measure any news that originated from citizens' reports to the MOE to avoid confounding both phenomena—although we alternatively use the full set of news as a robustness check. To avoid any type of researcher bias in the coding of these data, we outsourced the collection of this information to third parties. Finally, in using these data, we aggregate all of the irregularities in the news occurring in each municipality, but we also report the effects on specific types of irregularities. Appendix A provides further details about how we constructed this dataset.

To construct our second measure of electoral irregularities, we borrow from the electoral forensics literature (see Hicken and Mebane (2017) for a review), which uses data-driven methods to detect electoral irregularities. In particular, for each municipality, we test for deviations of the second digits of the vote counts at the voting booths from the 'natural' distributions these digits should follow in the absence of manipulation according to this literature. In section 4.2.2, we discuss the specific tests we use and their interpretation. In constructing these tests we use the official voting records provided by the *Registraduría Nacional del Estado Civil*, Colombia's electoral office, at the voting booth level.

3. Electoral outcomes. We also examine whether our intervention affected local elections. Specifically, we study whether it affected turnout, the vote share of particular candidates, or electoral competition. We compute all of these outcomes at the municipality level from the official voting records held by the *Registraduría Nacional del Estado Civil*.

3.5.2 Pre-Treatment Survey

We conducted a pre-treatment survey to gather additional information that was not available from existing external sources. We conducted this survey between October 7, three weeks before election day, until October 21, two days before the advertisement campaign began (see Figure 2). We recruited participants for the survey through Facebook advertisements that targeted users in the municipalities in the sample and invited them to participate in the survey.[21] As with the main campaign, we displayed the survey recruitment ads to all users of ages 18 or older in the sample of municipalities. Take-up was incentivized by including those who completed the survey in a raffle for several Samsung tablets (valued at $120 USD).

The survey took approximately 10-15 minutes to answer, and the recruitment ad did not refer to its content nor about the upcoming elections to avoid biased responses.[22] Our final sample is made of 6,121 complete responses coming from 630 municipalities, so the average number of responses per municipality is approximately 10. The main goal of this survey was to collect data on voters' perceptions about the mayoral candidates running in their municipalities, which we used in the analyses that we discuss later in detail in Section 4.3. Our final sample is balanced across treatment conditions in terms of survey respondents' sociodemographics (see Table A3), as well as in terms of the characteristics of the municipalities from which we obtained responses (see Table A9).

3.5.3 Other Covariates

We collected a rich set of municipal-level covariates to conduct balance checks and to include as controls in the main specifications. We mentioned these variables in Section 3.3 when we described the balance checks performed. In addition, we also collected candidate-level covariates, such as their sex, age, party and type of electoral platform (i.e, single party, coalition of parties, or independent). In Section B of the Appendix we describe all these variables in more detail and indicate their sources.

3.6 Empirical Analysis

Our main specification to estimate the average treatment effect of the different interventions is the following:

$$y_m = \mathbf{T}'_m \beta + X'_m \gamma + \epsilon_m \tag{1}$$

where y_m is the outcome variable for municipality m; $\mathbf{T}_m$ is a vector of indicators for the different treatment arms; X_m is a set of municipal covariates, including a set of fixed effects for the strata used in the randomization; and ϵ_m is the error term.

[21] Using Facebook ads as a survey recruitment strategy has been previously studied in the literature in both developed and developing contexts. It has been shown to be particularly effective at reaching populations that are costly to reach through conventional survey methods (Samuels and Zucco, 2013), such as the one at hand, and to approximate the representativeness of common recruitment methods such as phone surveys (Zhang et al., 2018).

[22] The main header in these ads read "Your opinion counts! Take our survey and participate in the raffle of three Samsung tablets. It will not take more than 10 minutes".

We estimate equation (1) using four different partitions of the treatment arms to study the effects of different sub treatments.[23] First, at the coarser level, we take $\mathbf{T}_m$ to simply include an indicator variable that takes the value of one if municipality m is in any of the treatment arms receiving ads and zero if it is in the placebo control group. Second, we study the differences between the specific messages featured in the ads by including separate indicators for whether municipality m received the information message, the call-to-action message or both, in vector $\mathbf{T}_m$. Third, we analyze the effect of the letter sent to politicians by including indicators for whether the municipality m received any of the reporting ads and either (i) we did not send a letter to candidates, or (ii) we did send it. Finally, we study possible differential effects between the full knowledge and partial knowledge letters sent to candidates, by including separate indicators for whether municipality m was included in any of these letter treatment arms, or it received any of the treatment arms but candidates did not receive any letter.

We report Huber-White standard errors for estimates of equation (1), along with randomization inference p-values to allow for inference that does not depend on distributional assumptions or asymptotic theory (Athey and Imbens, 2017; Young, 2017).

Instead of specifying the covariates to be included in X_m, we use the double-post-lasso covariate selection method proposed by Chernozhukov et al. (2015) and Belloni et al. (2014) to choose them. This method increases the efficiency of the estimates without running into overfitting issues. Crucially, it will include covariates for which there is unbalance across treatment arms as well as important correlates of the outcome variables considered. Unless otherwise specified, the set of covariates in Panels A-D in Table A2 along with the set of strata fixed effects is the one used when relying on this method. As detailed in Section 4.5, we also report estimates without control variables as a robustness check.

As discussed in Section 3.4, while the number of users viewing Facebook advertisements was balanced across treatment arms, the ads did not uniformly reach citizens across individual municipalities in our sample. These differences will distort the size of the treatment effects because they create differences in the intensity of treatment across municipalities. To correct this, we weight observations by the percentage of the target population (i.e., those of ages 18 years and above) who saw any of the ads. We use the number of users viewing the ads in each municipality, as provided by Facebook, as the numerator of these weights (i.e., the number of unique users who were displayed with an ad), and the municipalities' population estimates as the denominator.[24] As a robustness check we also report the estimates without

[23] This approach allows us to study the effect of different treatment conditions without sacrificing the statistical power needed to estimate the full model with nine different treatment arm combinations.

[24] Importantly, notice that this variable is balanced across all treatment arms (see Table A2).

these weights in Section 4.5.

4 Main Results

4.1 Results on Reporting

We begin by examining whether the reporting campaign was successful in inducing citizens to report irregularities and, in particular, high-quality reports with hard evidence that could be useful to prosecute offenders. Figure 5 provides a visual representation of the results of the estimation of equation (1) on reporting outcomes. The graph shows the estimated effects and, for reference, includes the control mean for each one of the outcomes. Overall, the evidence indicates that the reporting campaign increased reporting substantially, both on the extensive (i.e., the likelihood of reporting in a given municipality) as well as on the intensive margin (i.e., the number of reports filed).

Table 2 presents the full results of the estimations. First, looking at the results in Panel A, we find that receiving any of the treatments increased the probability that a report was filed from a municipality by 9.3 percentage points ($p < 0.05$), which corresponds to an increase of 33% compared to the control mean. Similarly, the number of reports increased in these municipalities by about 0.34, a 61% increase compared to the control.

Next, we examine if the intervention affected the subset of higher quality reports, as defined in Section 3.5.1. As shown in Table 2, municipalities receiving any of the treatments were 7 percentage points more likely to submit a report deemed as high-quality by the MOE than control ones ($p < 0.1$), and they increased the number of these types of reports by 0.17 ($p < 0.05$) - which represent, respectively, a 44% and 85% increase compared to the control group.

In Panel B of this table, we further explore *how* the ads' messages might have impacted citizens' incentives to report. To do so, we separately estimate the effect of each version of the ads. We start by pointing out that across outcomes, the ads containing the information of where individuals could report—i.e., the link to MOE's website—seem to be the ones driving the positive treatment effects. In particular, the 'call-to-action' message does not change citizen reporting in a statistically significant way, while the municipalities receiving the information message saw an increase in the probability of filing reports by 13-18 percentage points and the number of reports by 0.43-0.53 ($p < 0.05$), depending on the specification. Also note that there are no statistically significant differences between the 'information' and the 'information + call-to-action' messages, rejecting the hypothesis that the interaction of both versions could boost the incentives to report even further than the information message alone. In general, the ads providing information about the reporting website have strong effects on reporting at both the extensive and intensive margins.

Next, we study whether the letters sent to candidates had differential effects on reporting in Panel C. Interestingly, we find that although reporting increased in municipalities included in the campaign, both with and without the letter sent to candidates, the effects are larger when no letter is sent—although this difference is only significant at the extensive but not the intensive margin. We interpret this result in light of the results reported in the following section, which suggests that the letter sent to candidates might have had an *extra* deterrence effect over candidates' decision to engage in electoral irregularities, which in turn might have reduced reporting by citizens.

Finally, in Panel D we study whether the partial or full knowledge letters sent to candidates have any differential effects on reporting. Intuitively, we expect the full knowledge letters—which informed candidates about the specific website used in the campaign—to have the effect of increasing the number of reports more acutely, given that candidates themselves might submit reports in this condition. In line with this intuition, we find that full knowledge letters increased the overall number of reports as well as the high-quality reports filed ($p < 0.05$), while the partial knowledge letters did not. Furthermore, we find that the difference between these treatment arms is significant ($p < 0.1$)—although this does not hold in the extensive margin.

4.2 Effects on Electoral Irregularities

We now examine whether the intervention might have reduced the occurrence of electoral irregularities. To the extent that the reporting campaign was public, it is possible that campaign staffers and candidates were informed about it and that they changed their behavior in response to the threat of being reported. However, as mentioned in Section 3.1, candidates' knowledge about the reporting campaign and their subsequent reaction to it would have been substantially constrained by the fact that the advertisement campaign started only three days before election day—which would have given them relatively little time to both inform themselves about it and to react to this information. We thus expect that the letters sent to candidates informing about the reporting campaign—which were sent almost two months before election day—to generate a larger behavioral response on candidates and, in particular, a larger deterrence effect on irregularities.

Given the illicit nature of electoral irregularities, measuring them has been one of the main challenges in the literature. A popular strategy has been to use citizen reports (e.g., Rueda 2017) or survey responses to infer the occurrence of different types of irregularities (e.g., Blattman et al. 2019). However, in our context, this is not possible given that (i) reporting is directly affected by the intervention in potentially opposite ways to its effect on irregularities (i.e. the campaign increases reporting, as shown in the last section, but it might have reduced actual irregularities), and (ii) the reporting campaign might have affected citizens' willingness to speak about irregularities without having affected their actual occurrence.

We develop two strategies to measure electoral irregularities that overcome these difficulties, which we describe in the following two sections.

4.2.1 Media-Based Measure of Electoral Irregularities

Our first measure of electoral irregularities comes from an original database of electoral irregularities covered by the media, described in detail in Section 3.5.1. We construct two outcome measures using this dataset: an indicator for whether any irregularity was mentioned in the news in a given municipality, and the number of such irregularities.

Figure 6 depicts the main treatment effects on these variables, while Table 3 shows the estimates of the different sub treatment arms. These results indicate that the campaign had a large and negative effect on electoral irregularities, at both the extensive and intensive margins. Receiving any of the treatments reduced the probability of irregularities occurring by 8 percentage points ($p < 0.05$), and the number of irregularities by 0.1 ($p < 0.05$). These effects are particularly large given that they represent reductions of approximately 57% and 29%, respectively, compared to the control group mean.

When examining the effect of the different ad messages (Panel B of Table 3), we observe that the ads containing information about the reporting website are the main driving force behind these treatment effects. While the estimates for the call-to-action message are also negative, they are not statistically significant—although the difference between these treatment arms is not statistically significant in most cases.

As expected given our previous discussion about the short period that candidates had to learn and react to the reporting campaign, we find that the group of municipalities where we sent letters to candidates experienced a larger estimated reduction in both the probability and number of irregularities (see Panel C of the same table). Specifically, the municipalities in which candidates received a letter were 9 percentage points less likely to have an irregularity covered by the media and experienced a reduction of 0.12 in the number of such irregularities ($p < 0.01$). In turn, those exposed to the reporting campaign but where we did not send the letter saw a reduction of only 6 percentage points in the likelihood and 0.07 in the number of reports—and these effects are not significant. Yet, in drawing these comparisons, it is important to notice that the difference between the municipalities receiving the letter and those that did not (but, in both cases, exposed to the reporting campaign) is not significant. Furthermore, we do not find any statistically or economically meaningful differences in the effects of letters with full or partial knowledge (see Panel D).

As explained in Section 3.5.1, we purge all news about irregularities that arise from citizen reports to the MOE from our main measure of electoral irregularities to avoid confounding reporting and actual irregularities. One potential problem in excluding report-related news is that we might be selectively ignoring (actual) irregularities in municipalities treated by the reporting campaign. As a robustness check, we then consider an alternative measure containing *all* irregularities mentioned in the news—whether they mention reports made to MOE or not—and we report the results in Table A4. We see that although the precision and magnitude of some of the estimated effects drop, the main results hold, and the estimates remain statistically significant.

In addition to considering the overall effect of the interventions on all types of electoral irregularities, we also examine the effect on particular types of irregularities in Table A5 in the Appendix. Although the estimates are noisy given that some types of irregularities have a substantially low probability of occurrence, we see that the reduction of irregularities seems to be concentrated on vote-buying, whose likelihood drop by almost 5 percentage points ($p < 0.05$) in municipalities exposed to the reporting campaign, and on riots related to the elections, which decrease by 3 percentage points (not significant using conventional standard errors, but significant at a 10% level when using randomization inference p-values). Despite the fact that these two types of irregularities seem to be the main drivers of the effect, most types of irregularities also experience a decrease in their probability of occurrence. In fact, the main results remain unaltered when excluding types of irregularities one at a time from our main measure (see Table A6).

4.2.2 Forensic Measure of Electoral Irregularities

Our second measure of electoral irregularities comes from the election forensics literature, which uses "anomalies" in the administrative voting data to infer the occurrence of irregularities (see Hicken and Mebane (2017) for a review). In particular, we test for deviations of the second digit of voting counts from Benford's 2nd Digit Law—a "natural" distribution of second digits—as a proxy for the occurrence of electoral irregularities.

Although this type of test remains one of the most popular tools in the forensic literature,[25] some have cautioned against interpreting it as a *sufficient* proof of electoral irregularities (Deckert et al., 2011; Mebane, 2011). In our setting, these concerns are alleviated by the fact that the we will focus on comparing the relative adherence to Benford's Law of municipalities in our treatment arms as proxies of electoral irregularities, instead of considering

[25] A related 1st Digit Law has also been proposed, but the evidence indicates that it is not suited to detect fraud in contexts, such as ours, in which the maximum number of vote counts per voting booth are capped at relatively low numbers (300 in our setting) since these caps distort the natural distribution of the first digit (Pericchi and Torres, 2011).

this test as an absolute fail/pass measure of the occurrence of irregularities. Moreover, we use this measure as *complementary* to our news-based measure in assessing the occurrence of electoral irregularities—as has been advised by Hicken and Mebane (2017). It complements our news-based measure by detecting less conspicuous electoral irregularities, which might not be covered by the news but still leave a trace in administrative records. As shown in Table A7 in the Appendix, both types of measures are positively and significantly correlated ($\rho \approx 0.1 - 0.2$), but this correlation is not perfect, as expected from this discussion.

We use three different widely-used tests to verify compliance with Benford's 2nd Digit Law: Pearson's χ^2 test, the Kolmogorov-Smirnov test, and the Kuiper test. We describe the respective test statistics in Appendix C. While the Pearson χ^2 test is probably the most commonly used in the literature, it has been shown to be under-powered in small samples (Nigrini, 2012), such as the municipalities in our sample, which have typically 100-200 observations to compute these tests. The latter two tests are more appropriate for these types of samples and, in particular, the Kuiper test takes into account the "circular" nature of the distribution of second-digits. To further correct for small sample inference, we compute p-values simulating draws from the distribution under the null hypothesis that the data come from a Benford distribution.

For ease of interpretation, we synthesize the results of these tests into three main outcome variables. First, we construct a standardized index of all three test statistics, so that larger values reflect larger deviations from Benford's 2nd Digit Law. And second and third, we use an indicator variable which takes the value of one if any of the tests rejects the null hypothesis —i.e., there is statistical evidence of deviations—with p-value< 0.1 or p-value< 0.05, respectively.

Figure 7 presents a visual representation of the estimated effects of the interventions on these three outcomes, while Table 4 reports the point estimates for the full set of sub treatments. Across measures, we find that the municipalities that were exposed to the reporting campaign experienced fewer deviations from Benford's distribution, suggesting that less electoral irregularities took place in these locations. In general, rejection of the null hypothesis was substantial in the study sample, with 60% of the municipalities rejecting it at 10% significance level, and 52% rejecting it at a 5% significance level across any of the tests. Being exposed to any of the interventions in the reporting campaign reduced this substantially, by 11 and 12 percentage points respectively ($p < 0.05$), while, at the intensive margin, the tests statistics decreased by 0.17 standard deviations ($p < 0.05$).

Consistent with our findings using the media-based irregularity measure, we also find that (i) municipalities receiving the information message had a larger drop in the deviation from Benford's distribution than the ones receiving the call-to-action message (although this difference is not statistically significant), and (ii) that the letter sent to candidates had an "extra" effect in terms of the magnitude of the reduction in the deviation from Benford's 2nd Digit Law (although we do not find this effect is statistically different from the effect on the group exposed to the reporting campaign but where there were no letters sent to candidates).

Overall, the evidence in Table 4, combined with the one discussed in the previous section indicates that the reporting campaign not only increased citizens' engagement in monitoring elections by reporting irregularities, but it also generated a deterrence effect on candidates' engagement in electoral irregularities.

As a robustness check, in Table A8 we report the results when using each of the three considered tests separately. Consistent with the fact that the Pearson χ^2 test is less powered to reject the null hypothesis in small samples, only 28% of municipalities in the control group reject the null under this test with a significance level of 5% or less, while 42% and 34% of them do so when using the Kolmogorov-Smirnov and the Kuiper test, respectively. As seen in this table, the main results remain qualitatively unchanged by the choice of tests, although the effects are not statistically significant in the case of the Pearson χ^2 test, while they are significant and larger in the case of the two remaining tests.

4.3 Effects on Election Outcomes

The evidence presented in the last sections shows that the reporting campaign not only increased citizen reporting but also had a robust and substantial effect deterring electoral irregularities. In so far as some candidates' electoral prospects might have depended on engaging in irregularities (e.g., the votes they might have bought or the voters they might have intimidated to vote for them), we may expect the reporting campaign to also reduce their vote share. In this section, we begin by studying this possibility and then examine if the intervention affected other electoral outcomes, such as turnout. In the next section, we discuss alternative channels through which the reporting campaign might have affected electoral outcomes and also provide estimates of how much of the changes can be accounted for by the observed reduction in electoral irregularities.

Identifying the Candidates Likely to Engage in Irregularities. While the reporting campaign might have depressed the vote share for candidates whose success depended on electoral irregularities, it might not have affected other candidates. In our setting, the main challenge in studying these differences is identifying which candidates could have been involved in irregularities had our intervention not occurred. The latter is especially difficult given the large number of candidates running in the mayoral elections in each municipality and the scarce information about most candidates' backgrounds.[26]

In order to overcome this difficulty, we use citizens' responses in the pre-treatment survey we described in Section 3.5.2 to identify which candidates were more likely to engage in electoral irregularities in each municipality. In particular, we asked respondents to state whether each candidate would engage in different types of electoral irregularities. We did this for the most common irregularities described in Section 2—namely vote-buying, illicit advertisement, campaigning by public servants, voter intimidation, fraud in voter registration, and electoral fraud.

We then aggregated respondents' answers to these questions to construct three different variables at the candidate level. First, we compute the percentage of respondents that state that a candidate was going to engage in any electoral irregularity. Denote this first variable as Z^1_{mc}, for candidate c in municipality m. Second, we created a within-municipality version of this variable by subtracting the municipality's mean from Z^1_{mc}, so that positive values indicate that the candidate is more likely than the average one to engage in electoral irregularities. Thus, if $\bar{Z}_c$ is the municipality's mean, our second variable is defined as $Z^2_{mc} = Z^1_{mc} - \bar{Z}_c$. Finally, we also created an indicator of whether Z^1_{mc} was above average, so that $Z^3_{mc} = \mathbb{1}\left\{Z^1_{mc} \geq \bar{Z}_c\right\}$. While the first variable captures the full candidate-level variation in the likelihood to engage in irregularities that we aim to test, the two latter variables have the advantage of netting out municipal-level factors that could influence responses in ways unrelated to candidates' behavior.[27]

A potential concern in using these survey-based measures is that citizen views about candidates might not be good indicators about the actual behavior of candidates in general—and, in particular, when it comes to hard-to-observe behavior such as engagement in electoral misbehavior. In order to validate our measures vis-à-vis this concern, we contrast them with an external measure that does not rely on citizen input for its construction. In particular, we take advantage of the fact that the *Fundación Paz y Reconciliación* (PARES), a well-known NGO in Colombia, conducted a background check of the candidates in 48 of the municipalities in our sample and constructed a dataset indicating if each candidate had a history of past malfeasance, such as involvement in corruption or association with violent organizations.[28]

Despite the fact that the sample of municipalities investigated by PARES is relatively

[26] Upon registering their candidacy, candidates are subject to a legal background check by the government. Thus, there is virtually no variation in candidates' legal history, and differences in their likelihood to engage in electoral irregularities has to be captured in some other way.

[27] For instance, in "pessimistic" municipalities, most respondents might say that candidates will engage in irregularities regardless of whether they indeed will. The second and third measures will alleviate this concern by focusing on the within-municipality variation.

[28] The full report and dataset can be accessed here.

small and is not representative of the municipalities in our sample,[29] this outcome allows us to study how this "objective" measure compares to our survey-based methods. As reported in Table A10 in the Appendix, being found to have a history of malfeasance by PARES is strongly and significantly correlated with all of our survey-based measures and, in particular, with the within-municipality measures, which have correlations of ≈ 0.4.

Candidate-Level Regressions. We use our candidate-specific measures of the the likelihood of engaging in electoral irregularities to estimate the following regressions at the candidate level:

$$\text{Vote Share}_{cm} = \mathbf{T}'_m\beta + \psi \mathrm{Z}_{cm} + \mathrm{Z}_{cm} \times \mathbf{T}'_m\delta + X'_{cm}\gamma + \epsilon_{cm} \tag{2}$$

where Vote Share_{cm} is the vote share obtained by candidate c running in municipality m; $\mathbf{T}_m$ is a vector including different indicators for partitions of the treatment groups as before; Z_{cm} is one of the candidate-specific measures of the likelihood of engagement in electoral irregularities; and X_{cm} are a set of municipal and candidate covariates. Given our previous discussion, we expect that $\delta < 0$, which would suggest that candidates more engaged in electoral irregularities would experience a drop in their vote share.

Following our randomization strategy, we cluster the standard errors at the municipal level when estimating these candidate level regressions. However, these standard errors fail to incorporate the uncertainty generated by the fact that our survey-based measures (Z_{cm}) are estimates themselves.[30] To account for this extra uncertainty, we report the p-values from a two-step clustered bootstrap procedure, which incorporates the variation coming from estimating Z_{cm} and then the equation (2). We give a detailed description of this procedure in Appendix D.

Given that the pre-treatment survey responses came from a subset of municipalities in the original sample, we recalculate the balance checks for these regressions and report the results in Table A9 in the Appendix. Results in this table show that the municipalities continue to be well balanced in this sub-sample across both municipal and candidate covariates. As before, we use the double-post-lasso covariate selection method proposed by Chernozhukov et al. (2015) and Belloni et al. (2014) to choose the municipal and candidate level covariates to be included in the regressions with the aim of increasing the efficiency of the estimates and to correct for the few remaining imbalances.

Results from the Candidate-Level Regressions. Table 5 presents the results of estimating

[29] In conversations with representatives of PARES they explained that the main selection criteria were to choose municipalities where there were high risks to electoral transparency due to a history of violence, corruption or mining interests.

[30] An issue known as "generated regressors" in two-step estimations (Murphy and Topel, 1985).

equation (2) using the three different survey-based measures of candidates' likelihood to engage in irregularities. Across measures, we find that candidates more likely to engage in electoral irregularities experienced a large decrease in their vote share when they were running in municipalities included in the reporting campaign interventions. The magnitudes are quite substantial and significant: the reporting campaign reduced the vote share of candidates for which 100% of respondents say they will be engaged in electoral irregularities by over 4.2 percentage points ($p < 0.05$), by 6.7 for the demeaned version of this variable ($p < 0.01$), and by 3.8 for candidates above the municipal average according to this variable. Notice that, although large, these estimates are in the 'ballpark' of other interventions to curb electoral irregularities.[31]

Mimicking the results from the electoral irregularity outcomes, we see that the effect on these candidates' vote share is largest (i) for municipalities receiving the information message, and (ii) for municipalities where the candidates received a letter (although this only applies to the specifications using the within-municipality measures).

In the last column of Table 5 we alternatively estimate equation (2) using an indicator variable that takes the value of one if the candidate was found to have a history of past malfeasance according to the investigation conducted by PARES. Despite this regression being limited to the municipalities that they investigated, we find estimates consistent with our results using the survey-based measures. In particular, we find that the reporting campaign reduced the vote share of candidates with histories of past malfeasance by 7.7 percentage points ($p < 0.1$).

As a robustness check, we verify whether, rather than measuring the propensity to engage in electoral irregularities, our measures capture candidates' popularity or visibility in the weeks leading to the election.[32] To test this hypothesis, we estimated equation (2) interacting different measures of the popularity of candidates with the treatment indicators. In particular, we use two different proxies of candidates' popularity: (i) the proportion of respondents in the pre-treatment survey who say that the candidate will win the election, and (ii) an indicator for whether the candidate is running with an incumbent party or not. The results, presented in Table A12 in the Appendix, show that the heterogeneity of the treatment conditions with these measures of popularity does not produce significant estimates of δ—nor is the magnitude of these estimates as large as the one using the heterogeneity

[31]For instance, in the context of Russia, Enikolopov et al. (2013) find that polling stations where electoral observers were assigned experienced a decrease of 11 p.p. in the vote share for the incumbent party. Similarly, Blattman et al. (2019) find that an anti-vote-buying campaign in Uganda decreased the vote share of incumbents by approximately 3 p.p.

[32]As reported in Table A11 in the Appendix, the tested measures are indeed correlated with the popularity of candidates—measured as either the proportion of respondents who say that the candidate will win the election or whether the candidate is running for an incumbent party.

involving the propensity of candidates to engage in electoral irregularities.

Additional Electoral Outcomes. We now examine if the reporting campaign also impacted other electoral outcomes, such as turnout or the margin of victory. In particular, understanding whether the drop in the vote share for candidates likely to engage in irregularities came from a change in turnout might help understand the channels through which the reporting campaign generated this effect. Similarly, the intervention could have also affected the level of political competition, measured as the margin of victory of the elected mayor.

Figure 8 displays the main effects of the intervention over these two outcomes by estimating equation (1), while Table 6 reports the point estimates for the full set of sub treatments. Overall, we find that the intervention did not affect neither of these outcomes. In particular, the fact that turnout was not significantly affected suggests hat the decrease in the vote share for candidates more likely to engage in irregularities must have come with a parallel increase in the vote share for "cleaner" candidates.

4.4 Interpreting the Effects on Electoral Outcomes

The evidence presented in the previous section shows that the reporting campaign had the effect of reducing the vote share for candidates likely to engage in electoral irregularities. As argued so far, this effect might have been a consequence of the decline in electoral irregularities occurring in municipalities exposed to the reporting campaign interventions that might have, in tandem, reduced the vote share of the candidates relying on these electoral irregularities to boost their electoral prospects.

However, there is at least one other channel that might mediate in explaining these results. In particular, it is plausible that the reporting campaign could have raised the salience of the issue of electoral irregularities to voters by exposing them to this topic through Facebook advertisements. In turn, this might have changed their voting decisions by voting against the candidates they perceived to be at the source of the issue of electoral irregularities—which would be consistent with our findings.

In practice, it is possible that both the drop in electoral irregularities and the increased salience of irregularities determined the overall effect reported in the previous section.[33] One feature of our experimental design that allows us to take a first step in figuring the relative contribution of both channels in determining the overall effect is the presence of the "call-to-action" message treatment condition. Municipalities in this group would have been exposed to the increased salience about irregularities, but, as we have seen in the

[33] A third channel which might have affected the vote share of candidates is the substitution of resources used from electoral irregularities to legal campaigning efforts. However, we do not emphasize this channel since it is likely that it might have dampened the drop in the vote share of candidates more likely to engage in irregularities, rather than to explain the effect we found.

previous sections, they experienced no significant changes in reporting nor in the occurrence of electoral irregularities. Thus, they provide a group in which most (if not all) of the changes in electoral outcomes would have come from the salience channel and not the electoral irregularity channel. As seen in Table 5, the effects on the vote share of candidates more likely to engage in electoral irregularities are not statistically significant and are between 26% and 70% smaller than the ones from the group receiving the information message. Assuming the effect of the call-to-action treatment captures the full extent of the salience channel, this would imply that between 26% and 66% of the total decrease in the vote share of candidates likely to engage in irregularities is due to the electoral irregularity channel.

As an alternative way of quantifying the contribution of this channel, we formalize the previous discussion with a simple model. Suppose the vote share obtained by candidate c, conditional on her likelihood of engaging in irregularities (Z_c) is given by the following function $V(\cdot)$:

$$V_c = V(I(T), O(T); Z_c)$$

where $I(T)$ are the electoral irregularities that candidate c engages in, given the treatment status T; $O(T)$ are other factors influencing the vote share of the candidate c –including the salience of irregularities—which depend on T. Given this expression, the treatment effect on the vote share of c can be decomposed as:

$$\frac{\mathrm{d}V_c}{\mathrm{d}T} = \frac{\mathrm{d}V_c}{\mathrm{d}I}\frac{\mathrm{d}I}{\mathrm{d}T} + \frac{\mathrm{d}V_c}{\mathrm{d}O}\frac{\mathrm{d}O}{\mathrm{d}T} \tag{3}$$

We are interested in the first term of equation (3): the effect of irregularities on candidate c's vote share, which depends on treatment assignment T. We have direct estimates of $\frac{\mathrm{d}I}{\mathrm{d}T}$ from the results presented in the previous sections. On the other hand, we do not have causal estimates for $\frac{\mathrm{d}V_c}{\mathrm{d}I}$, but we can approximate this by using correlations in the control group.[34] Using these estimates, we can then approximate the percentage of the total effect on vote share due to the change in electoral irregularities by computing $\frac{\mathrm{d}V_c}{\mathrm{d}T} \times 100/\left(\frac{\mathrm{d}V_c}{\mathrm{d}I}\frac{\mathrm{d}I}{\mathrm{d}T}\right)$.[35]

[34]In practice we get an estimate for $\frac{\mathrm{d}V_c}{\mathrm{d}I}$ by estimating coefficient α_3 in the following regression:

$$\text{Vote Share}_{cm} = \alpha_1 I_m + \alpha_2 \mathrm{Z}_{cm} + \alpha_3 \mathrm{Z}_{cm} \times I_m + u_{cm}$$

using different proxies for irregularities (I_m) and Z_{mc}.

[35]To give an example of this procedure, consider using the rejection of Benford's 2nd Digit Law at a 95% significance as a proxy for I_m and the the vote share of candidates for which the percentage of respondents that say they will engage in electoral irregularities is above average as Z_c. From column (3) in Table 4 we have $\frac{\mathrm{d}I}{\mathrm{d}T} \approx -0.12$. For the correlation in the control data we have $\frac{\mathrm{d}V_c}{\mathrm{d}I} \approx 8.56$. Thus, $\frac{\mathrm{d}V_c}{\mathrm{d}I} \times \frac{\mathrm{d}I}{\mathrm{d}T} \approx -1.03$, which would imply that 26.6% of the overall drop in the vote share reported in Table A17 is due to the electoral irregularity channel.

In Table A13 in the Appendix we report the results of performing this exercise using different combinations of the proxies used for electoral irregularities (I_m) and Z_c in the previous analysis. The results range between 11% and 35%, which are similar to estimates comparing the effects of the call-to-action and information message treatment arms. While this exercise is far from perfect, it provides complementary evidence of the relative importance of the electoral irregularity channel in determining the drop in the vote share of candidates likely to engage in electoral irregularities, as we report in the last section.

4.5 Additional Robustness

In this section, we show that our results are robust to different methodological decisions in estimating the treatment effects of the interventions. In particular, we study the robustness with respect to (1) not including any controls in the estimation, (2) not weighting the observations by the percentage of ad viewers in the municipal population.

Tables A14-A18 present the results of running all analyses without controls except for fixed effects for the strata used in randomization. As expected, we see that across outcomes the magnitude of the treatment effects is virtually unaffected, but that they are less precisely estimated. Despite this, we see that most of the treatment effects discussed in the previous sections remain significant and the main conclusions remain unaltered.

In Tables A19-A23 we rerun all analyses without using weights. Once again, we see that the sign and statistical significance of our estimators remains broadly unchanged across outcomes compared to our main specifications. However we see that the magnitude of some estimates does change slightly—which is expected given that these specifications weight equally municipalities with no ad viewers (and thus the effect of the intervention should had been null) and municipalities with large numbers of viewers.

5 Cost-Benefit Analysis

The effectiveness of the reporting campaign is better appreciated in terms of its cost-effectiveness compared to other similar strategies to reduce electoral irregularities that have been studied in the literature. While estimating the cost-effectiveness of our intervention is straightforward, comparing it to other strategies is complicated since (i) there are relatively few papers reporting the costs of their interventions, and (ii) many papers do not have estimates of the effect of these interventions on irregularities, and even when they do, there are no homogeneous measurements of irregularities. In the following, we limit ourselves to comparisons to the few papers that report costs and we compare their cost-effectiveness as measured by their impact on votes for candidates benefiting from irregularities—which is a measure provided by most papers. Following our discussion in the introduction, we organize our comparison exercise by focusing on (i) interventions studying the use of electoral observers, (ii) the use of top-down ICT technologies to monitor elections, and (iii) the use of voter-education campaigns against irregularities.

Our two main benchmark metrics for the cost-effectiveness of our intervention are the following. First, a back of the envelope calculation suggests that our intervention reduced 1 percentage point of the vote share of candidates above average in the percentage of people who say they will engage in electoral irregularities at the cost of \$0.45 USD per polling station.[36] Second, a similar calculation suggests that a single dollar spent in the reporting campaign reduced 36 votes for these same candidates.[37]

1. Comparison to electoral observers. Despite being the strategy most used and studied to curb electoral irregularities, virtually no paper studying the effects of electoral observers reports their cost. In a discussion of their cost-effectiveness, Callen et al. (2016) estimate that the European Union spends about $\$6,000 - \$20,000$ USD per polling station deploying electoral observers in developing countries. These large costs undermine the cost-effectiveness of this strategy, even if we consider the most "optimistic" evaluations of their impact. For instance, Enikolopov et al. (2013) find that the random deployment of electoral observers around elections in Russia generated a drop of 11 percentage points per polling station of the vote share of the allegedly corrupt incumbent party—which is the largest effect reported for these types of interventions that we know of. Taking the range of costs per polling station mentioned before, this would imply that a 1 percentage point change in the vote share costs between \$545 and \$1818 USD per polling station, which is several orders of magnitude higher than the costs of our intervention.[38]

2. Comparison to top-down ICT monitoring. Two papers report the cost-effectiveness of ICT monitoring interventions in reducing electoral irregularities. Both of them study the impact of announcing the implementation of a campaign to take pictures of electoral tallies before aggregating vote counts, as a way of reducing aggregation fraud. In the first paper, Callen and Long (2015) report that deploying this intervention in Afghanistan costs

[36]The Facebook advertisements cost $\$10,870$ USD in total. This implies a cost of \$15.57 USD per municipality or \$1.71 USD per polling station (since there $6,349$ polling stations in the municipalities in our sample). Given that the effect of the reporting campaign was to reduce the vote share of candidates above average in the percentage of people who say they will engage in electoral irregularities by 3.801 p.p., the cost of reducing a single percentage point is \$0.45 USD ($= 1.71/3.801$) per polling station.

[37]The average votes in each municipality are 13352, which means that the reporting campaign reduced by 507.51 ($= 3.801/100 \times 13,352$) the votes for candidates above average in the percentage of people who say they will engage in electoral irregularities. This then implies that a single dollar invested in the campaign was responsible for a reduction of 32.60 votes (507.51/15.57) per municipality.

[38]These estimates might be overstated given that the costs reported by Callen et al. (2016) refer to less developed countries than Russia. However, even if real costs were 10% of the lower considered ones, this intervention would still be more than two orders of magnitude costlier than the one considered in this paper.

approximately \$210 USD per polling station[39] and reduced by 6 the votes for candidates "connected" to electoral authorities (which are the ones most able to benefit from fraud). This would imply a cost of \$35 USD per vote. In the second paper, Callen et al. (2016) report a cost of \$40 USD per polling station in Uganda, which reduced the vote share of the incumbent candidate by 3 percentage points,[40] which implies that reducing a single percentage point cost approximately \$13 USD. The intervention considered in both of these papers is substantially more cost-effective than deploying electoral observers but is still considerably less so than the reporting campaign we study. This is mainly because implementing this strategy requires sending staff to take pictures of voting tallies. Crowd-sourcing this task to decrease these costs, in a spirit similar to our intervention, might be a promising avenue for future research.

3. Comparison to voter-education interventions. The lion's share of the voter education campaigns studied in the literature involves the training and mobilization of staff to convey messages and develop activities related to reducing electoral irregularities (Blattman et al., 2019; Collier and Vicente, 2013; Hicken et al., 2018; Vasudevan, 2019; Vicente, 2014). Although we can legitimately suspect such efforts probably involve large costs relative to their reported effects on irregularities, virtually none of these papers report the costs of their studied interventions. One exception is Vasudevan (2019), who, instead of relying on in-person campaigns, explored the effect of a radio campaign informing citizens of the economic consequences of vote-buying in India. He finds that the intervention was extremely cost-effective in reducing the votes of candidates coming from vote-buying parties, with a one-dollar investment translating into 78 fewer votes for these candidates—approximately twice as cost-effective as the reporting campaign we deployed. Importantly, however, the experimental design of this intervention ruled out possible reactions of candidates, which could have possibly muted its effect.[41] This implies that the practical implementation of this intervention, which would entail the full knowledge and reaction of candidates, would have led to smaller cost-effectiveness estimates. However, the fact that this intervention produces estimates in the same order of magnitudes as our considered intervention highlights the great promise of using media to enhance electoral integrity.

[39] They report a "[...]total budget of just over US\$100,000" for 471 treated polling stations, which are the numbers used to construct this estimate.

[40] This is the largest estimate reported by them.

[41] Vasudevan (2019) argues that since the radio campaign occurred in the three days before the elections when electioneering is prohibited and most vote-buying occurs in this context, candidates would not have the chance to react to it.

6 Conclusion

Despite substantial efforts by both governments and international agencies to fight electoral irregularities, these remain an important issue that hampers accountability and development in a large part of the world. In this paper, we provide evidence that ICT technologies aimed at incorporating civil society in the oversight of elections are an effective way to promote electoral integrity in the context of widespread malpractice.

The policy implications of these findings are three-fold. First, we find that using a social media campaign to induce citizens to report is an effective way to spur high-quality, evidence-backed reports that can be used by competent authorities to prosecute the authors of electoral irregularities. Moreover, we find this is the case even in a context in which there are pre-existing and popular reporting websites. This highlights the importance of providing information and reminders to citizens about where and how to report through widely disseminated outlets such as social media ads. Second, we find that this intervention deters electoral electoral irregularities in a more cost-effective way than other strategies traditionally used by governments, NGOs, and international organizations —e.g., election observers and different education campaigns. Finally, by relying on social media to disseminate our reporting campaign, we believe this type of strategy can easily be scaled up in contexts where there is enough internet and social media penetration. Moreover, given the rapid expansion of the internet throughout the developing world, these types of strategies will be increasingly more suitable for the most remote regions, where alternative strategies would be particularly more costly and government accountability is worse (World Bank, 2017).

The findings in this paper also open several important avenues for future research. To begin with, how do these types of interventions affect citizens' views and trust about democracy and the government more generally? Although we do not find effects on turnout, access to voice mechanisms and increased accountability might generate positive changes in citizen attitudes toward democratic institutions, but this probably requires longer exposure to the mechanisms to emerge. Second, what are the downstream effects of this type of intervention on the accountability of governments and the provision of public goods and services? By decreasing the popular support of candidates more engaged in electoral irregularities, these interventions might generate positive effects on the selection and agency of candidates. Third, we have shown the potential of a social network like Facebook to amplify the scope and effect of campaigns that encourage citizen monitoring of elections. However, our experimental design does not allow us to fully test for their efficacy compared to other dissemination vehicles (e.g., leaflets or traditional media).

References

Acemoglu, Daron, James Robinson, and Rafael J. Santos-Villagran (2013). "The Monopoly of Violence: Evidence from Colombia". *Journal of the European Economic Association* 11, pp. 5–44.

Arenas, Natalia (2018). "El primer eslabón: La compra de los ediles". In: *El dulce poder: Así funciona la política en Colombia*. Penguin Random House, pp. 51–58.

Asunka, Joseph, Sarah Brierley, Miriam Golden, Eric Kramon, and George Ofosu (2019). "Electoral Fraud or Violence: The Effect of Observers on Party Manipulation Strategies". *British Journal of Political Science* 49(1), pp. 129–151.

Athey, Susan and Guido Imbens (2017). "The Econometrics of Randomized Experiments". In: *Handbook of Economic Field Experiments*. Ed. by Abhijit Banerjee and Esther Duflo. Vol. 1. North-Holland, pp. 73–140.

Baland, Jean-Marie and James A. Robinson (2007). "How Does Vote Buying Shape the Economy". In: *Elections for Sale: The Causes and Consequences of Vote Buying*. Ed. by Frederic Charles and Schaffer Andreas. Lynne Rienner Publishers.

Banerjee, Abhijit, Esther Duflo, Clement Imbert, Santhosh Mathew, and Rohini Pande (2020). "E-Governance, Accountability, and Leakage in Public Programs: Experimental Evidence froma Financial Management Reform in India". *American Economic Journal: Applied Economics* 12(4), pp. 39–72.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014). "Inference on Treatment Effects after Selection among High-Dimensional Controls". *The Review of Economic Studies* 81(2), pp. 608–50.

Blattman, Christopher, Horacio Larreguy, Benjamin Marx, and Otis Reid (2019). "Eat Widely, Vote Wisely? Lessons from a Campaign Against Vote Buying in Uganda". Working paper.

Callen, Michael, Clark C. Gibson, Danielle F. Jung, and James D. Long (2016). "Improving Electoral Integrity with Information and Communications Technology". *Journal of Experimental Political Science* 3, pp. 4–17.

Callen, Michael, Saad Gulzar, Ali Hasanain, Muhammad Yasir Khan, and Arman Rezaeek (2020). "Data and Policy Decisions: Experimental Evidence from Pakistan". *Journal of Development Economics* 146, pp. 1–10.

Callen, Michael and James D. Long (2015). "Institutional Corruption and Election Fraud: Evidence from a Field Experiment in Afghanistan". *American Economic Review* 105(1), pp. 354–381.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2008). "Bootstrap-Based Improvements For Inference With Clustered Errors". *The Review of Economics and Statistics* 90(3), pp. 414–427.

Chernozhukov, Victor, Christian Hansen, and Martin Spindler (2015). "Post-Selection and Post-Regularization Inference in Linear Models with Many Controls and Instruments". *American Economic Review: Papers and Proceedings* 105(5), pp. 486–490.

Collier, Paul and Pedro C. Vicente (2013). "Votes and violence: Evidence from a field experiment in Nigeria". *The Economic Journal* 124, pp. 327–355.

Deckert, Joseph, Mikhail Myagkov, and Peter C Ordeshook (2011). "Benford's Law and the Detection of Election Fraud". *Political Analysis* 19(3), pp. 245–268.

DellaVigna, Stefano and Matthew Gentzkow (2010). "Persuasion: Empirical Evidence". *Annual Review of Economics* 2(1), pp. 643–669.

Driscoll, Jesse and Daniel Hidalgo (2014). "Intended and Unintended Consequences of Democracy Promotion Assistance to Georgia After the Rose Revolution". *Research and Politics*, pp. 1–13.

Enikolopov, Ruben, Vasily Korovkin, Maria Petrova, Konstantin Sonin, and Alexei Zakharov (2013). "Field Experiment Estimate of Electoral Fraud in Russian Parliamentary Elections". *Proceedings of the National Academy of Sciences* 110(2), pp. 448–452.

Fergusson, Leopoldo, Carlos Molina, and Juan Felipe Riaño (2017). "I Sell My Vote, And So What? A New Database and Evidence From Colombia". *Documentos CEDE* (20), pp. 1–13.

Finan, Frederico and Laura Schechter (2012). "Vote-Buying and Reciprocity". *Econometrica* 80(2), pp. 863–881.

Fox, Johnathan A. (2015). "Social Accountability: What Does the Evidence Really Say?" *World Development* 72, pp. 346–361.

Garcia, Miguel and Sebastian Pantoja (2015). "Incidencia del clientelismo segun riesgo electoral y de violencia: Un análisis de las elecciones presidenciales de 2014 en municipios de consolidación territorial". In: *Mapas y factores de riesgo electoral. Elecciones de autoridades locales.* Mision de Observación Electoral, pp. 291–313.

Gonzalez, Robert (2021). "Cell Phone Access and Election Fraud: Evidence from a Spatial Regression Discontinuity Design in Afghanistan". *American Economic Journal: Applied Economics* 13 (2), pp. 1–51.

Gonzalez-Ocantos, Ezequiel, Chad Kiewiet de Jonge, Carlos Meléndez, Javier Osorio, and David W. Nickerson (2012). "Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua". *American Journal of Political Science* 56(1), pp. 202–217.

Hicken, Allen (2011). "Clientelism". *Annual Review of Political Science* 14, pp. 289–310.

Hicken, Allen, Stephen Leider, Nico Ravanilla, and Dean Yang (2018). "Temptation in Vote-Selling: Evidence from a Field Experiment in the Philippines". *Journal of Development Economics* 131, pp. 1–14.

Hicken, Allen and Walter R. Mebane (2017). "A Guide to Elections Forensics". *Research and Innovation Grants Working Papers Series: USAID*, pp. 1–38.

Holland, Alisha C and Brian Palmer-Rubin (2015). "Beyond the machine: Clientelist brokers and interest organizations in Latin America". *Comparative Political Studies* 48(9), pp. 1186–1223.

Hyde, Susan D. (2007). "The Observer Effect in International Politics: Evidence from a Natural Experiment". *World Politics* 60(1), pp. 37–63.

Hyde, Susan D. (2010). "Experimenting in Democracy Promotion: International Observers and the 2004 Presidential Elections in Indonesia". *Perspectives on Politics* 8(2), pp. 511–527.

Ichino, Nahomi and Matthias Schündeln (2012). "Deterring or Displacing Electoral Irregularities? Spillover Effects of Observers in a Randomized Field Experiment in Ghana". *The Journal of Politics* 74(1), pp. 292–307.

Kelley, Judith (2012). *Monitoring Democracy: When International Election Observation Works and Why it Often Fails*. Princeton University Press.

Khemani, Stuti (2015). "Buying Votes Versus Supplying Public Services: Political Incentives to Under-Invest in Pro-Poor Policies". *Journal of Development Economics* 117, pp. 84–93.

Leal Buitrago, Francisco and Andrés Dávila-Ladrón de Guevara (1990). *Clientelismo: El sistema politico de Colombia y su expresión regional*.

Leeffers, Stefan and Pedro C. Vicente (2019). “Does Electoral Observation Influence Electoral Results? Experimental Evidence for Domestic and International Observers in Mozambique”. *World Development* 114, pp. 42–58.

Lewis-Faupel, Sean, Yusuf Neggers, Benjamin A. Olken, and Rohini Pande (2016). “Can Electronic Procurement Improve Infrastructure Provision? Evidence from Public Works in India and Indonesia”. *American Economic Journal: Economic Policy 2016, 8(3): 258–283* 8(3), pp. 258–283.

Mebane, Walter R. (2011). “Comment on “Benford’s Law and the Detection of Election Fraud””. *Political Analysis* 19(3), pp. 269–272.

Misión de Observación Electoral (2018). *Irregularidades electorales en Colombia: Informe final Pilas con el voto, Elecciones presidencia y congreso 2018*.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar (2016). “Building State Capacity: Evidence from Biometric Smartcards in India”. *American Economic Review* 106(10), pp. 2895–2929.

Murphy, Kevin M. and Robert H. Topel (1985). “Estimation and Inference in Two-Step Econometric Models”. *Journal of Business & Economic Statistics* 3(4), pp. 370–379.

Nichter, Simeon (2008). “Vote Buying or Turnout Buying? Machine Politics and the Secret Ballot”. *American Political Science Review* 102(1), pp. 19–31.

Nigrini, Mark J. (2012). *Benford’s Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. Vol. 586. John Wiley & Sons.

Olken, Benjamin and Rohini Pande (2012). “Corruption in Developing Countries”. *Annual Review of Economics* 4, pp. 479–505.

Peixoto, Tiago and Jonathan Fox (2016). “When Does ICT- Enabled Citizen Voice Lead to Government Responsiveness?” *Background paper for the World Development Report 2016*, pp. 1–26.

Pericchi, Luis and David Torres (2011). “Quick Anomaly Detection by the Newcomb–Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela”. *Statistical Science* 26(4), pp. 502–516.

Robinson, James A. and Ragnar Torvik (2014). “The Real Swing Voter’s Curse”. *American Economic Review: Papers and Proceedings* 99(2), pp. 310–315.

Rueda, Miguel R. (2017). “Small Aggregates, Big Manipulation: Vote Buying Enforcement and Collective Monitoring”. *American Journal of Political Science* 61(1), pp. 163–177.

Samuels, David and Cesar Zucco (2013). "Using Facebook as a Subject Recruitment Tool for Survey-Experimental Research". SSRN Working paper.

Simpser, Alberto and Daniela Donno (2019). "Can International Election Monitoring harm Governance?" *Journal of Politics* 74(2), pp. 501–513.

Singer, Matthew M. (2009). "Buying Voters with Dirty Money: The Relationship between Clientelism and Corruption". Presented at the annual American Political Science Association Meeting.

Stokes, Susan C. (2005). "Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina". *American Political Science Review* 99(3), pp. 315–325.

Stokes, Susan C., Thad Dunning, Marcelo Nazareno, and Valeria Brusco (2013). *Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics*. Cambridge University Press.

Vasudevan, Srinivasan (2019). "Diminishing the Effectiveness of Vote Buying: Experimental Evidence from a Persuasive Radio Campaign in India". Working paper.

Vicente, Pedro C. (2014). "Is Vote-buying Effective? Evidence from a Field Experiment in West Africa". *The Economic Journal* 124(574), pp. 356–387.

Vicente, Pedro C. and Leonard Wantchekon (2009). "Clientelism and Vote Buying: Lessons from Field Experiments in African Elections". *Oxford Review of Economic Policy* 25(2), pp. 292–305.

World Bank (2016). *World Development Report: Digital Dividends*. Washington, DC: World Bank.

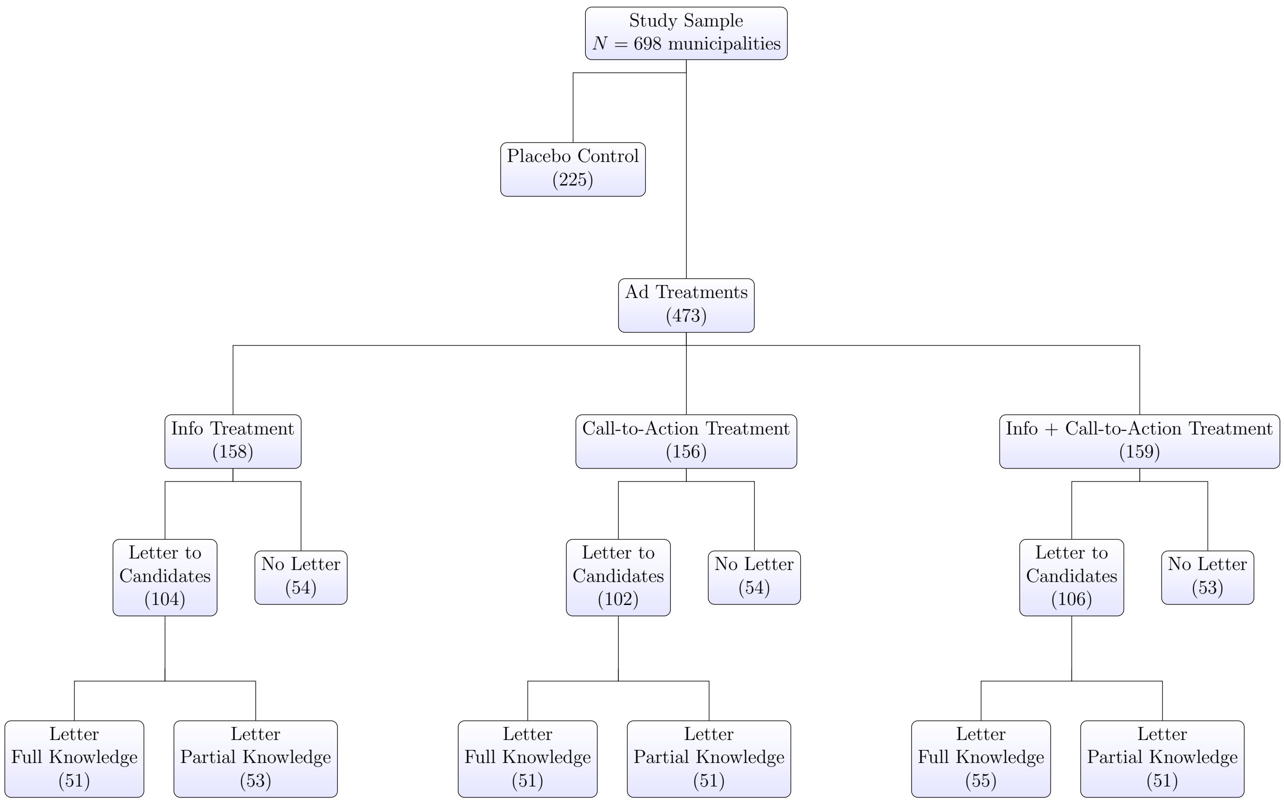
World Bank (2017). *World Development Report: Governance and the Law*. Washington, DC: World Bank.

Young, Alwyn (2017). "Channelling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results". London School of Economics.

Zhang, Baobao, Matto Mildenberger, Peter D. Howe, Jennifer Marlon, Seth A. Rosenthal, and Anthony Leiserowitz (2018). "Quota Sampling Using Facebook Advertisements". *Political Science Research and Methods*, pp. 1–7.

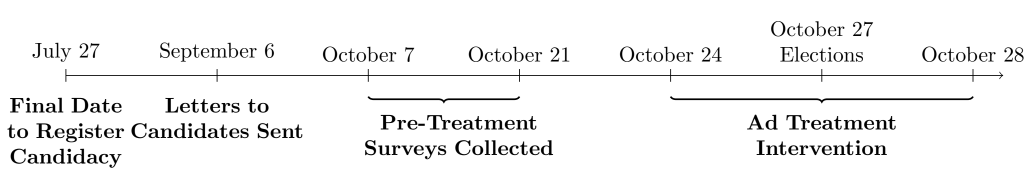
Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov (2020). "Political effects of the internet and social media". *Annual Review of Economics* 12, pp. 415–438.

Figure 1: Randomization Design



Notes: This figure illustrates the experimental design of the experiment. The sample size within each treatment group is shown in parenthesis.

Figure 2: Timeline of the Intervention



Notes: This figures shows the timeline of the interventions performed in the study. Note that the timeline is not drawn to scale.

Figure 3: Ad Slideshow - 2019 Experiment

(a) Slide A: *"Report Electoral Irregularities!"*

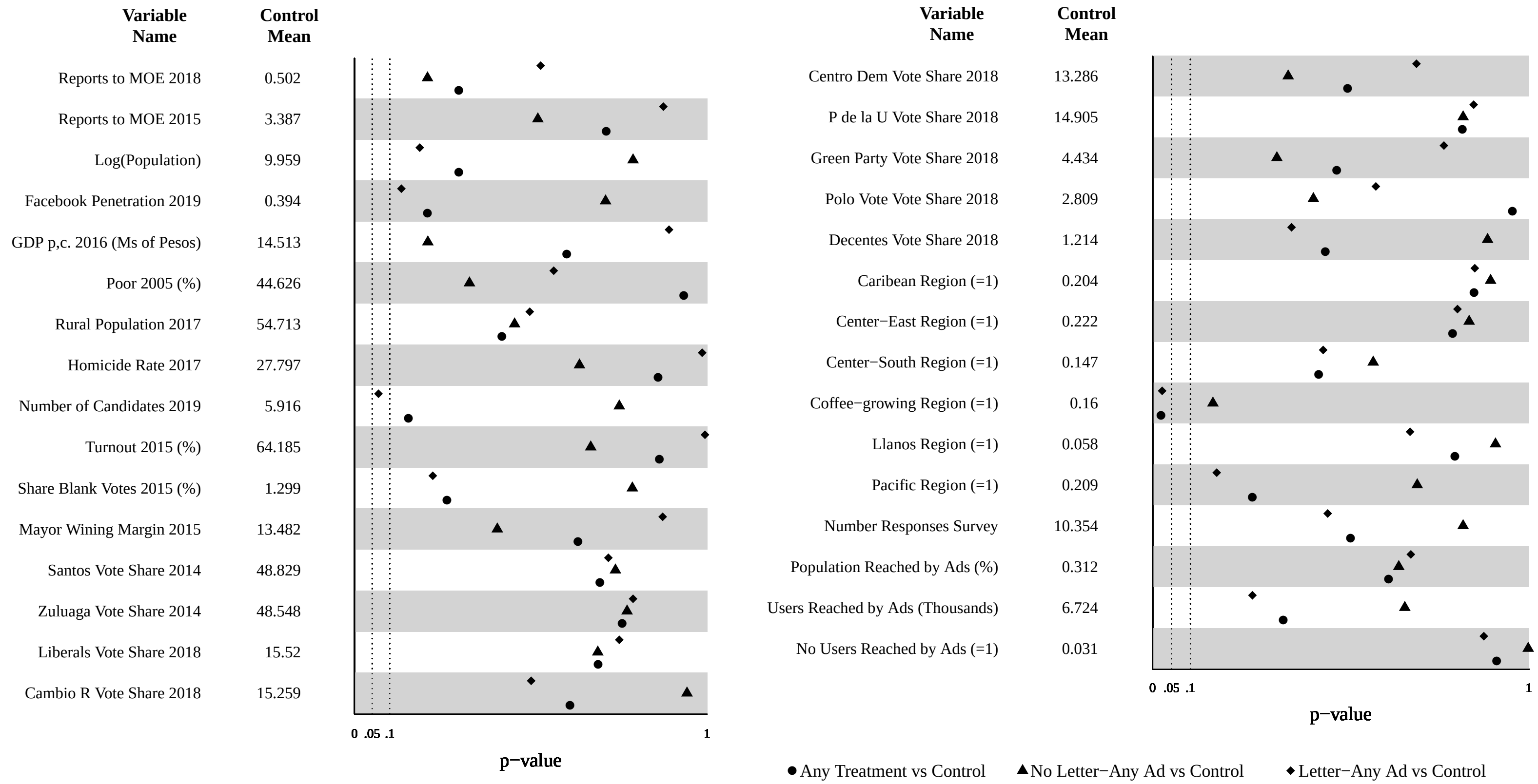
(b) Slide B: *"Reporting Website: Pilas con el voto"*

(c) Slide C: *"Sunday October 27"*

(d) Slide D: *"Next local elections"*

Notes: The four possible slides shown on the ad interventions in the second experiment are shown in this figure. Below each slide is a translation to English of the text contained in the slides. The Placebo Control group was shown only Slides C and D. The Information message group was shown slides B, C and D. The Call-to-Action message group was shown slides A, C and D. Finally, the group with both the Call-to-Action and the Information message was shown all of the slides, A-D.

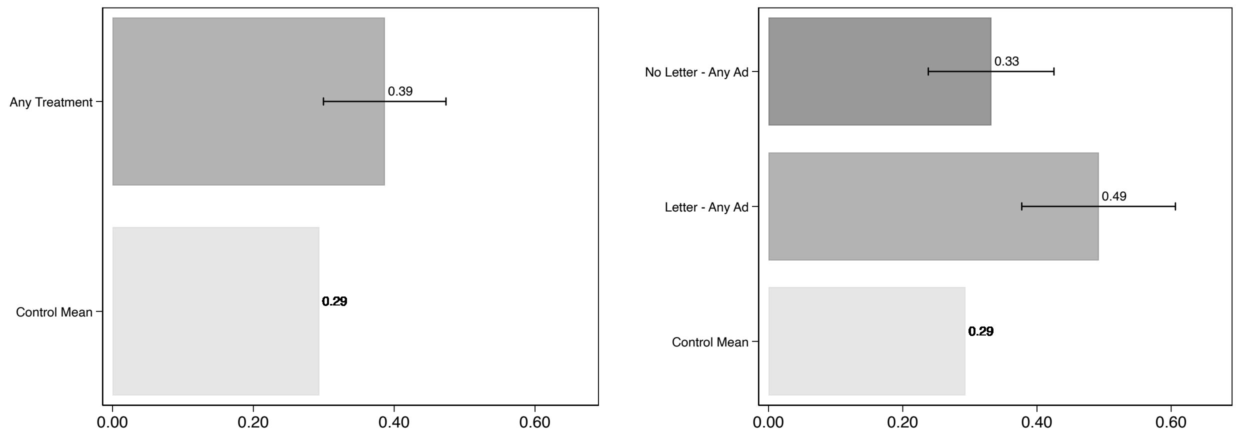
Figure 4: Covariate Balance



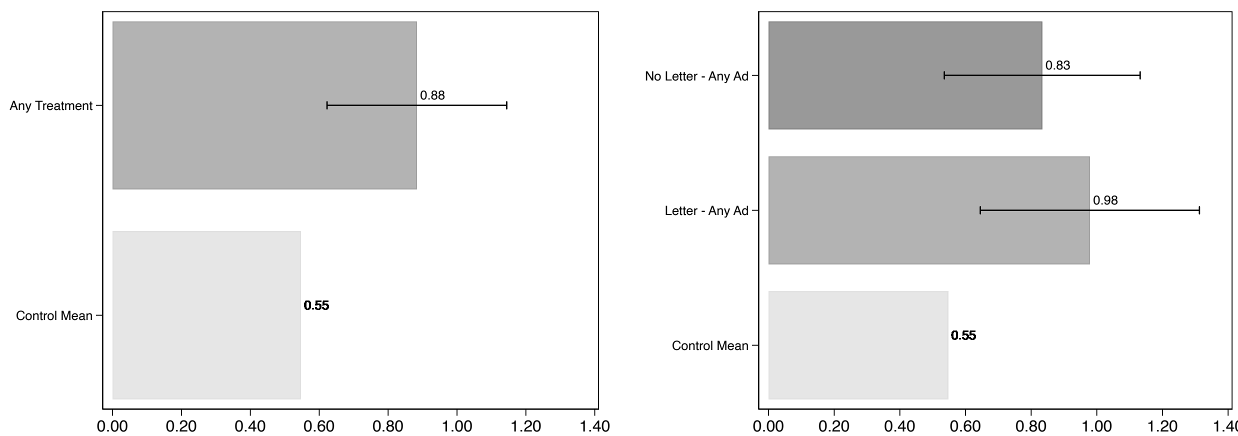
Notes: This table presents the p-values of the balance checks across three main treatment arms, using five sets of selected covariates: *Previous Reports:* reports to MOE 2018 and 2015; *Socioeconomic:* log of the population, Facebook penetration in 2019, GDP per capital 2016 (in Colombian pesos), percentage of poor in 2005, rural population in 2017, and homicide rate in 2017; *Political:* number of candidates registered in the 2019 mayoral elections, turnout in the 2015 mayoral elections, percentage of blank votes in 2015, the elected mayor's margin of victory in 2015, Santos' and Zuluaga's vote shares in the 2014 presidential elections, and the vote shares obtained by each party in the 2018 congressional elections; *Geographic:* regional dummies; and *Other:* number of responses in pre-treatment survey, percentage of population reached by the Facebook Ad, number of users reached by the Facebook Ad, and whether there were no users reached by the ad. The tests correspond to difference in means, in which observations are weighted by the percentage of the population older than 18 that was reached by a Facebook ad, except when reporting the difference in means of variables referring to the reach of the Facebook ads. The control group mean of each variable is presented.

Figure 5: Impacts on Reports

(a) Reports (=1)



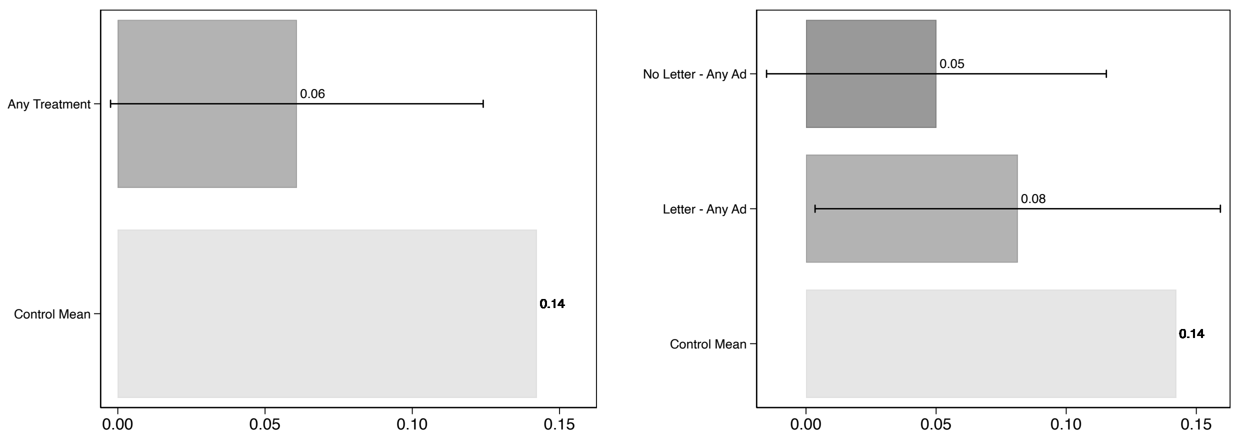
(b) Number of Reports



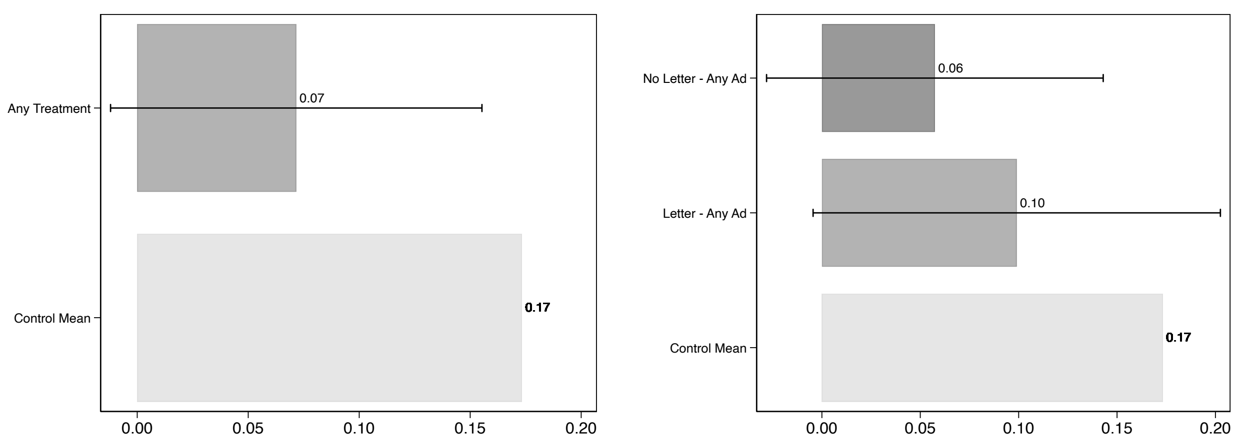
Notes: This figure reports the effects of the intervention on report outcomes. The outcome in panel (a) is an indicator for whether any report was issued to the MOE from each municipality. In panel (b) it is the number of such reports. The control group mean is shown in each plot, while the remaining bars report the control group mean plus the effect for each treatment arm estimated using specification (1). All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. 95% confidence intervals are reported.

Figure 6: Impacts on the Media-Based Irregularity Measures

(a) Media Irregularities (=1)



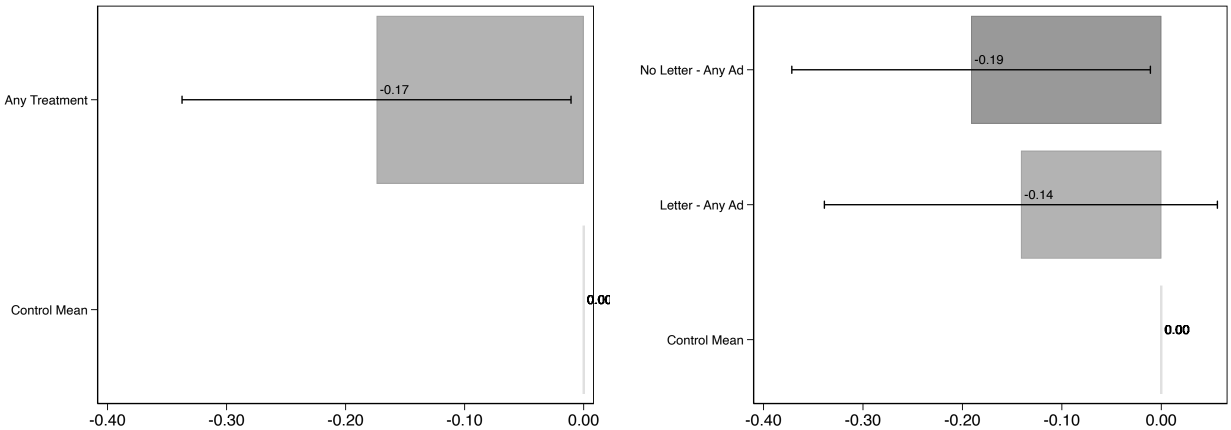
(b) Number of Media Irregularities



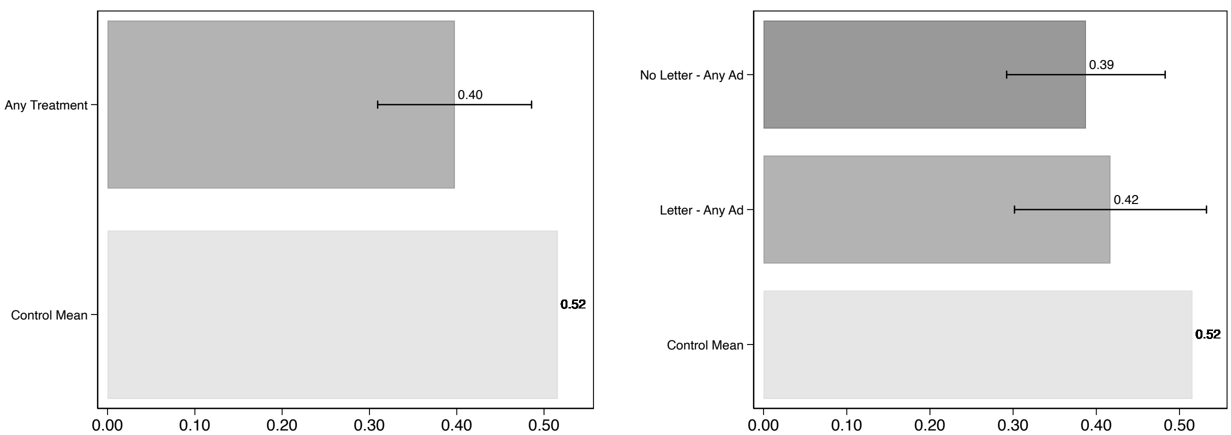
Notes: This figure reports the effects of the intervention on media-based electoral irregularity measures. The outcome in panel (a) is an indicator for whether any irregularity was reported in the media in a particular municipality. In panel (b) it is the number of different irregularities. The control group mean is shown in each plot, while the remaining bars report the control group mean plus the effect for each treatment arm estimated using specification (1). All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. 95% confidence intervals are reported.

Figure 7: Impacts on Deviations from Benford's Second Digit Law

(a) Index of all Forensic Test Stats (z-score)



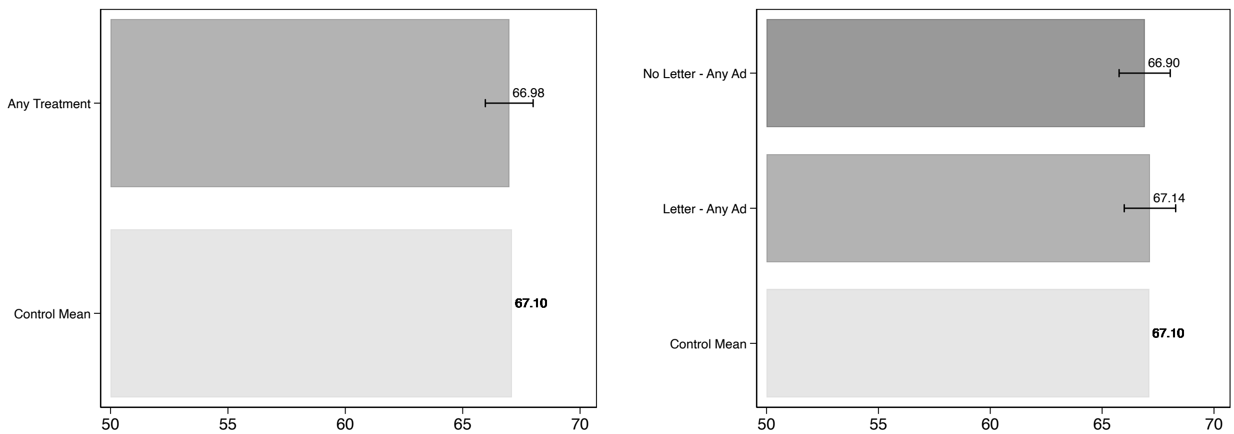
(b) Any P-value From Forensic Tests < 0.05 (=1)



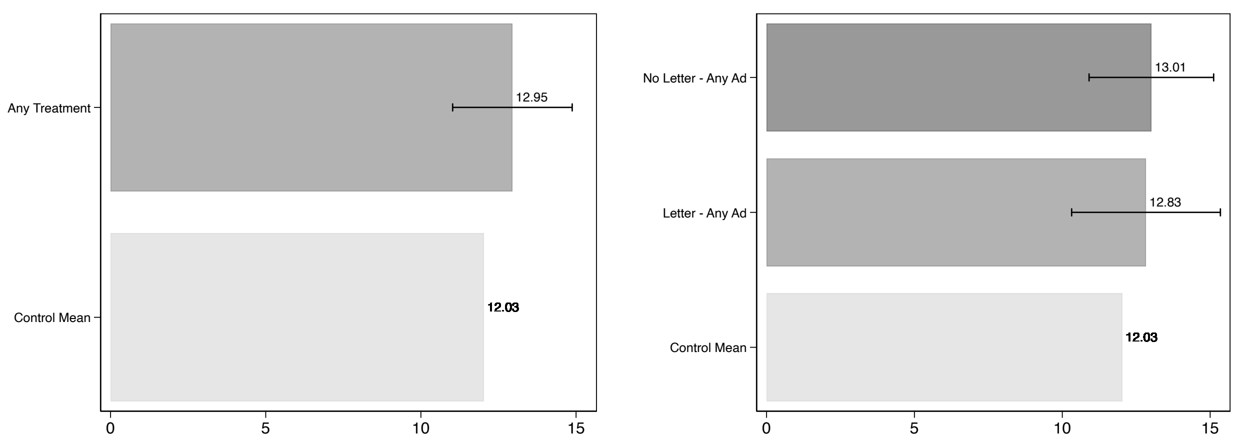
Notes: This figure reports the effects of the intervention on measures of the deviation from Benford's Second Digit Law . The outcome in panel (a) is the index of the χ^2, Kolmogorov-Smirnov and Kuiper test statistics testing for Benford's 2nd digit law, described in Section 4.2.2 . In panel (b) it is an indicator that takes the value of one if the p-value of any of these tests leads to rejection of the null hypothesis with less than a 5% significance level. The control group mean is shown in each plot, while the remaining bars report the control group mean plus the effect for each treatment arm estimated using specification (1). All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. 95% confidence intervals are reported.

Figure 8: Impacts on Additional Electoral Outcomes

(a) Turnout (%)



(b) Margin of Victory (%)



Notes: This figure reports the effects of the intervention on electoral outcomes . The outcome in panel (a) is turnout, as a percentage of the people registered to vote. In panel (b) it is the margin of the winning candidate over the runner-up , expressed as a percentage of total votes. The control group mean is shown in each plot, while the remaining bars report the control group mean plus the effect for each treatment arm estimated using specification (1). All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. 95% confidence intervals are reported.

Table 1: Scale of Ad Campaigns

	Total	Per municipality	Per Population > 18 years
Viewers of the Ad	4,358,870	6244.80	0.33
Times the Ad Appeared on a Screen	12,886,427	18,461.93	0.99
People Clicking on the MOE's Link*	23,418	33.55	1.79 (per 1000)
Users Reacting to Ad	9623	13.79	0.74 (per 1000)
Post Shares	4531	6.49	0.35 (per 1000)
Comments on Ad	437	0.63	0.03 (per 1000)

Notes: This table reports several measures of the scale of the Facebook advertisement campaign, as well as metrics of the engagement of users with the the ads. The variables reported in this table are defined as follows. 'Viewers of the Ad' are the number of distinct individuals who saw the ads at least once. 'Times the Ad Appeared on a Screen' are the number of times the ads appeared on any screen. 'People Clicking on the MOE's Link' are the number of distinct individuals who clicked on the link landing on the MOE's reporting website. $*$: For this metric, we only considered the municipalities in actually receiving the link to the MOE's website (i.e. the ones receiving the Information message) when computing the measures by municipality and population 18 or older. 'Users Reacting to Ad' are the number of distinct individuals who reacted to the ad by clicking on one of the available Facebook reactions (i.e. like, love, laugh, etc...). 'Post Shares' are the number of times people shared the ad in their own timeline, in other friends' timelines or in groups. 'Comments on Ad' are the number of comments made on the ads.

Table 2: Impacts on Reports

	(1) Reports (=1)	(2) N. Reports	(3) High Quality Reports (=1)	(4) High Quality N. Reports
Panel A. Pooled Treatment				
[T] Any treatment	0.093**	0.337**	0.071*	0.169**
	(0.044)	(0.133)	(0.038)	(0.066)
	[0.048]	[0.028]	[0.084]	[0.016]
Panel B. Subtreatments by Types of Ads				
[IA] Information Ad	0.129**	0.425**	0.051	0.096
	(0.058)	(0.189)	(0.048)	(0.077)
	[0.026]	[0.012]	[0.272]	[0.158]
[CA] Call-to-Action Ad	-0.028	0.050	0.053	0.117
	(0.057)	(0.166)	(0.050)	(0.095)
	[0.666]	[0.762]	[0.270]	[0.182]
[$I+CA$] Info + Call-to-Action Ad	0.176***	0.529***	0.109**	0.296***
	(0.058)	(0.191)	(0.051)	(0.113)
	[0.006]	[0.006]	[0.034]	[0.004]
Test $IA = CA$, p-value	0.01	0.07	0.97	0.84
Test $IA = I + CA$, p-value	0.46	0.65	0.30	0.10
Test $CA = I + CA$, p-value	0.00	0.02	0.33	0.18
Panel C. Subtreatments by Letter - No Letter				
[NL] No Letter - Any Ad	0.199***	0.432**	0.103**	0.198**
	(0.058)	(0.170)	(0.050)	(0.096)
	[0.000]	[0.004]	[0.040]	[0.018]
[L] Letter - Any Ad	0.038	0.287*	0.055	0.154**
	(0.048)	(0.152)	(0.041)	(0.075)
	[0.406]	[0.054]	[0.204]	[0.044]
Test $NL = L$, p-value	0.00	0.42	0.31	0.67
Panel D. Subtreatments by Types of Letters				
[NL] No Letter - Any Ad	0.199***	0.432**	0.104**	0.198**
	(0.058)	(0.170)	(0.050)	(0.096)
	[0.004]	[0.006]	[0.024]	[0.010]
[PL] Partial Knowledge Letter - Any Ad	0.001	0.078	0.028	0.020
	(0.057)	(0.163)	(0.048)	(0.066)
	[0.986]	[0.596]	[0.566]	[0.770]
[FL] Full Knowledge Letter - Any Ad	0.076	0.496**	0.082	0.289**
	(0.058)	(0.214)	(0.051)	(0.117)
	[0.212]	[0.004]	[0.096]	[0.002]
Test $NL = PL$, p-value	0.00	0.06	0.16	0.07
Test $NL = FL$, p-value	0.06	0.79	0.70	0.51
Test $PL = FL$, p-value	0.24	0.07	0.33	0.02
Control Mean	0.29	0.55	0.16	0.20
Sample Size	677	677	677	677

Notes: The outcome in column (1) is an indicator for whether any report was issued to the MOE from each municipality. In column (2) it is the number of such reports. In columns (3)-(4) the same definitions are used on the subset of reports of a high quality (see Section 2 for a discussion about how quality of reports is assessed by the MOE). All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table 3: Impacts on the Media-Based Irregularity Measures

	(1) Media Irregularities (=1)	(2) Number of Media Irregularities
Panel A. Pooled Treatment		
[T] Any treatment	-0.081**	-0.102**
	(0.032)	(0.043)
	[0.002]	[0.010]
Panel B. Subtreatments by Types of Ads		
[IA] Information Ad	-0.069*	-0.098**
	(0.040)	(0.049)
	[0.114]	[0.064]
[CA] Call-to-Action Ad	-0.057	-0.070
	(0.041)	(0.056)
	[0.176]	[0.214]
[$I + CA$] Info + Call-to-Action Ad	-0.117***	-0.137***
	(0.032)	(0.044)
	[0.000]	[0.004]
Test $IA = CA$, p-value	0.77	0.58
Test $IA = I + CA$, p-value	0.12	0.28
Test $CA = I + CA$, p-value	0.07	0.14
Panel C. Subtreatments by Letter - No Letter		
[NL] No Letter - Any Ad	-0.061	-0.074
	(0.040)	(0.053)
	[0.118]	[0.188]
[L] Letter - Any Ad	-0.092***	-0.116***
	(0.033)	(0.044)
	[0.004]	[0.000]
Test $NL = L$, p-value	0.33	0.31
Panel D. Subtreatments by Types of Letters		
[NL] No Letter - Any Ad	-0.061	-0.074
	(0.040)	(0.053)
	[0.136]	[0.196]
[PL] Partial Knowledge Letter - Any Ad	-0.081**	-0.116**
	(0.039)	(0.047)
	[0.040]	[0.018]
[FL] Full Knowledge Letter - Any Ad	-0.104***	-0.116**
	(0.035)	(0.049)
	[0.010]	[0.024]
Test $NL = PL$, p-value	0.61	0.35
Test $NL = FL$, p-value	0.20	0.38
Test $PL = FL$, p-value	0.48	0.99
Control Mean	0.14	0.17
Sample Size	677	677

Notes: The outcome in column (1) is an indicator for whether any irregularity was reported in the media in a particular municipality. In column (2) it is the number of different irregularities. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table 4: Impacts on Deviations from Benford's Second Digit Law

	(1) Index of all Forensic Test Stats (z-score)	(2) Any P-value From Forensic Tests < 0.1 (=1)	(3) Any P-value From Forensic Tests < 0.05 (=1)
Panel A. Pooled Treatment			
[T] Any treatment	-0.174**	-0.107**	-0.118***
	(0.083)	(0.045)	(0.045)
	[0.038]	[0.020]	[0.012]
Panel B. Subtreatments by Types of Ads			
[IA] Information Ad	-0.298***	-0.154***	-0.158***
	(0.105)	(0.059)	(0.058)
	[0.002]	[0.002]	[0.006]
[CA] Call-to-Action Ad	-0.086	-0.058	-0.072
	(0.103)	(0.059)	(0.059)
	[0.412]	[0.334]	[0.208]
[$I+CA$] Info + Call-to-Action Ad	-0.136	-0.107*	-0.123**
	(0.110)	(0.059)	(0.059)
	[0.184]	[0.066]	[0.046]
Test $IA = CA$, p-value	0.06	0.14	0.18
Test $IA = I + CA$, p-value	0.15	0.48	0.59
Test $CA = I + CA$, p-value	0.66	0.45	0.43
Panel C. Subtreatments by Letter - No Letter			
[NL] No Letter - Any Ad	-0.141	-0.069	-0.099*
	(0.101)	(0.059)	(0.059)
	[0.170]	[0.238]	[0.098]
[L] Letter - Any Ad	-0.191**	-0.126***	-0.128***
	(0.092)	(0.049)	(0.049)
	[0.038]	[0.008]	[0.008]
Test $NL = L$, p-value	0.60	0.31	0.60
Panel D. Subtreatments by Types of Letters			
[NL] No Letter - Any Ad	-0.141	-0.069	-0.099*
	(0.101)	(0.059)	(0.059)
	[0.182]	[0.242]	[0.108]
[PL] Partial Knowledge Letter - Any Ad	-0.252***	-0.144**	-0.139**
	(0.097)	(0.060)	(0.059)
	[0.004]	[0.022]	[0.008]
[FL] Full Knowledge Letter - Any Ad	-0.129	-0.109*	-0.117**
	(0.118)	(0.058)	(0.057)
	[0.200]	[0.052]	[0.036]
Test $NL = PL$, p-value	0.27	0.26	0.54
Test $NL = FL$, p-value	0.92	0.53	0.77
Test $PL = FL$, p-value	0.28	0.60	0.74
Control Mean	0.00	0.60	0.52
Sample Size	677	677	677

Notes: The outcome in column (1) is the index of the χ^2, Kolmogorov-Smirnov and Kuiper test statistics testing for Benford's 2nd digit law, described in Section 4.2.2 . In columns (2) and (3) it is an indicator that takes the value of one if the p-value of any of these tests leads to rejection of the null hypothesis with less than a 10% or 5% significance level. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table 5: Impacts on Vote Share of Candidates Likely to Engage in Irregularities

	(1)	(2)	(3)	(4)
	Vote Share (%)			
Interaction term Z:	Candidate will engage in irregularities (fraction of respondents)	Demeaned Candidate will engage in irregularities (fraction of respondents)	Above Average Candidate will engage in irregularities (=1)	Past Malfeasance (=1)
Panel A. Pooled Treatment				
$[T \times Z]$ Any treatment $\times Z$	-4.213**	-6.686***	-3.801***	-7.699*
	(2.009)	(2.433)	(1.296)	(4.333)
	[0.030]	[0.000]	[0.002]	[0.040]
	{0.037}	{0.007}	{0.005}	
Panel B. Subtreatments by Types of Ads				
$[IA \times Z]$ Information Ad $\times Z$	-4.703*	-7.581**	-3.800**	-11.843***
	(2.423)	(2.937)	(1.664)	(3.951)
	[0.056]	[0.026]	[0.024]	[0.022]
	{0.053}	{0.010}	{0.026}	
$[CA \times Z]$ Call-to-Action Ad $\times Z$	-1.445	-3.408	-2.811*	-4.697
	(2.569)	(3.256)	(1.694)	(8.110)
	[0.584]	[0.352]	[0.102]	[0.550]
	{0.588}	{0.304}	{0.097}	
$[I + CA \times Z]$ Info + Call-to-Action Ad $\times Z$	-6.165**	-8.806**	-4.866***	-2.048
	(2.657)	(3.438)	(1.686)	(5.157)
	[0.014]	[0.012]	[0.002]	[0.792]
	{0.024}	{0.011}	{0.005}	
Test $IA \times Z = CA \times Z$, p-value	0.22	0.23	0.59	0.36
Test $IA \times Z = I + CA \times Z$, p-value	0.59	0.73	0.56	0.02
Test $CA \times Z = I + CA \times Z$, p-value	0.10	0.16	0.27	0.75
Panel C. Subtreatments by Letter - No Letter				
$[NL \times Z]$ No Letter - Any Ad $\times Z$	-4.686**	-6.699**	-3.199*	-5.494
	(2.332)	(2.962)	(1.684)	(4.767)
	[0.060]	[0.024]	[0.062]	[0.364]
	{0.051}	{0.029}	{0.064}	
$[L \times Z]$ Letter - Any Ad $\times Z$	-3.959*	-6.680**	-4.115***	-9.449*
	(2.241)	(2.745)	(1.404)	(5.175)
	[0.068]	[0.014]	[0.004]	[0.128]
	{0.075}	{0.015}	{0.003}	
Test $NL \times Z = L \times Z$, p-value	0.74	0.99	0.57	0.45
Panel D. Subtreatments by Types of Letters				
$[NL \times Z]$ No Letter - Any Ad $\times Z$	-4.682**	-6.699**	-3.200*	-5.380
	(2.333)	(2.963)	(1.685)	(4.776)
	[0.050]	[0.040]	[0.058]	[0.388]
	{0.051}	{0.030}	{0.064}	
$[FL \times Z]$ Full Knowledge Letter - Any Ad $\times Z$	-3.434	-6.861**	-3.757**	-3.459
	(2.748)	(3.239)	(1.696)	(7.887)
	[0.158]	[0.052]	[0.030]	[0.606]
	{0.213}	{0.035}	{0.029}	
$[PL \times Z]$ Partial Knowledge Letter - Any Ad $\times Z$	-4.460*	-6.472*	-4.460***	-13.386**
	(2.640)	(3.536)	(1.681)	(5.440)
	[0.080]	[0.050]	[0.006]	[0.050]
	{0.095}	{0.068}	{0.007}	
Test $NL \times Z = PL \times Z$, p-value	0.93	0.95	0.50	0.15
Test $NL \times Z = FL \times Z$, p-value	0.65	0.96	0.77	0.81
Test $PL \times Z = FL \times Z$, p-value	0.73	0.92	0.71	0.24
Control Mean	18.91	18.91	18.91	18.91
Sample Size	2986	2986	2986	263
N. Municipalities	628	628	628	48

Notes: The outcome in all columns is the vote share for each candidate, expressed as a percentage of total valid votes. In each of these columns, a different measure of the likelihood that a candidate commits irregularities is used to compute the candidate-level heterogenous effects. In column (1) it is the proportion of respondents from the pre-treatment survey that say that the candidate might commit at least one type of electoral irregularity. In column (2) the outcome is this same variable, demeaned using the municipality-level mean. In column (3) it is an indicator that takes the value of one if this variable is above the municipal-level mean. Finally, in column (4) it is an indicator for whether a candidate was found to be involved in malfeasance in the past according to the investigation by the NGO PARES. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Clustered standard errors at the municipality level are shown in parentheses, random inference p-values are shown in square brackets, and clustered wild-bootstrap p-values correcting for the variance in estimating Z are shown in curly brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

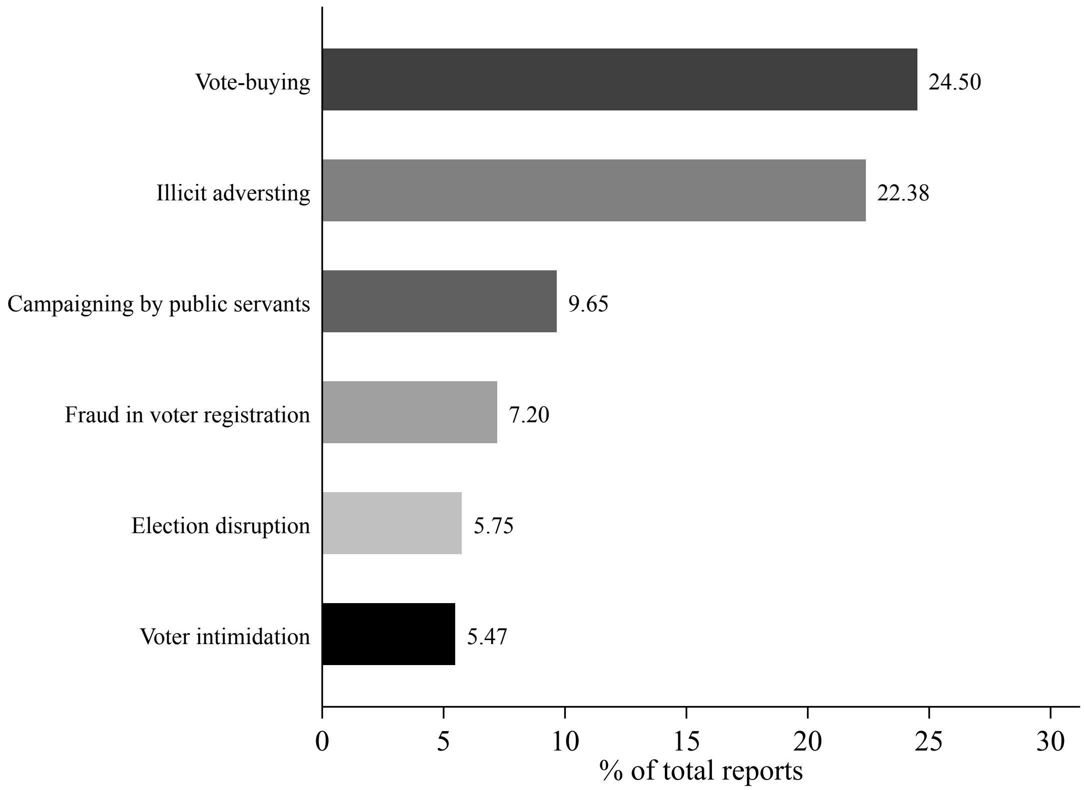
Table 6: Impacts on Additional Electoral Outcomes

	(1) Turnout (%)	(2) Margin of Victory (%)
Panel A. Pooled Treatment		
[T] Any treatment	-0.114	0.921
	(0.519)	(0.983)
	[0.824]	[0.346]
Panel B. Subtreatments by Types of Ads		
[IA] Information Ad	-0.129	1.130
	(0.666)	(1.427)
	[0.872]	[0.416]
[CA] Call-to-Action Ad	0.366	1.503
	(0.556)	(1.235)
	[0.544]	[0.256]
[$I+CA$] Info + Call-to-Action Ad	-0.570	0.120
	(0.770)	(1.253)
	[0.470]	[0.920]
Test $IA = CA$, p-value	0.43	0.81
Test $IA = I + CA$, p-value	0.59	0.52
Test $CA = I + CA$, p-value	0.21	0.31
Panel C. Subtreatments by Letter - No Letter		
[NL] No Letter - Any Ad	0.040	0.802
	(0.586)	(1.283)
	[0.942]	[0.556]
[L] Letter - Any Ad	-0.196	0.983
	(0.583)	(1.075)
	[0.764]	[0.424]
Test $NL = L$, p-value	0.67	0.89
Panel D. Subtreatments by Types of Letters		
[NL] No Letter - Any Ad	0.040	0.803
	(0.587)	(1.284)
	[0.944]	[0.514]
[PL] Partial Knowledge Letter - Any Ad	-0.309	0.165
	(0.685)	(1.228)
	[0.678]	[0.898]
[FL] Full Knowledge Letter - Any Ad	-0.084	1.807
	(0.731)	(1.390)
	[0.924]	[0.184]
Test $NL = PL$, p-value	0.61	0.65
Test $NL = FL$, p-value	0.86	0.51
Test $PL = FL$, p-value	0.78	0.28
Control Mean	67.10	12.03
Sample Size	677	677

Notes: The outcome in column (1) is turnout, as a percentage of the people registered to vote. In column (2) it is the margin of the winning candidate over the runner-up , expressed as a percentage of total votes. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

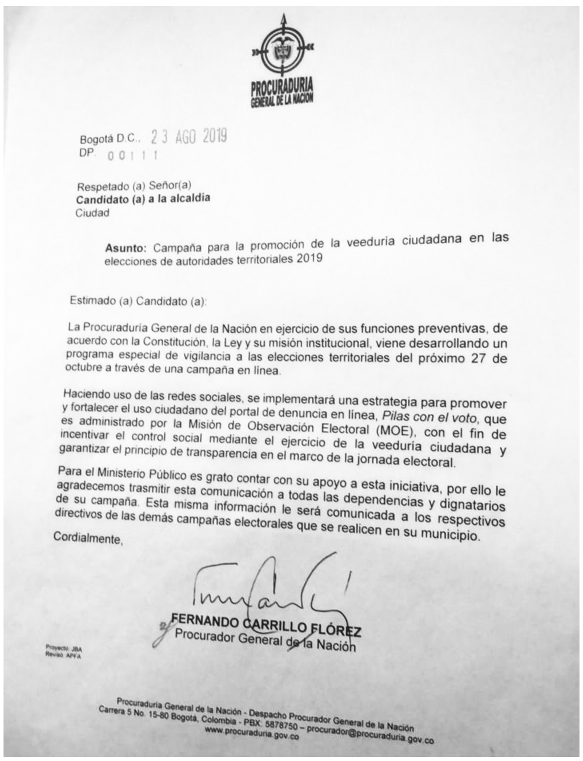
7 Appendix

Figure A1: Electoral irregularities Reported in 2015 Elections to the MOE



Notes: This figure displays the proportion of electoral irregularities of different types as a percentage of total irregularities reported through the MOE's *Pilas con el voto* in the context of the 2015 mayoral elections. The reports are restricted to those received between October 22-26, 2015 (the election day was on October 25). The definitions for each type of electoral irregularity are presented in Section 2.

Figure A2: Letter sent to candidates - Full Knowledge



PROCURADURIA
GENERAL DE LA NACION

Bogotá D.C., 23 AGO 2019
DP. 00111

Respetado (a) Señor(a)
Candidato (a) a la alcaldia
Ciudad

Asunto: Campaña para la promoción de la veeduría ciudadana en las elecciones de autoridades territoriales 2019

Estimado (a) Candidato (a):

La Procuraduría General de la Nación en ejercicio de sus funciones preventivas, de acuerdo con la Constitución, la Ley y su misión institucional, viene desarrollando un programa especial de vigilancia a las elecciones territoriales del próximo 27 de octubre a través de una campaña en línea.

Haciendo uso de las redes sociales, se implementará una estrategia para promover y fortalecer el uso ciudadano del portal de denuncia en línea, *Pilas con el voto*, que es administrado por la Misión de Observación Electoral (MOE), con el fin de incentivar el control social mediante el ejercicio de la veeduría ciudadana y garantizar el principio de transparencia en el marco de la jornada electoral.

Para el Ministerio Público es grato contar con su apoyo a esta iniciativa, por ello le agradecemos trasmitir esta comunicación a todas las dependencias y dignatarios de su campaña. Esta misma información le será comunicada a los respectivos directivos de las demás campañas electorales que se realicen en su municipio.

Cordialmente,

FERNANDO CARRILLO FLÓREZ
Procurador General de la Nación

Proyectó: JBA
Revisó: APFA

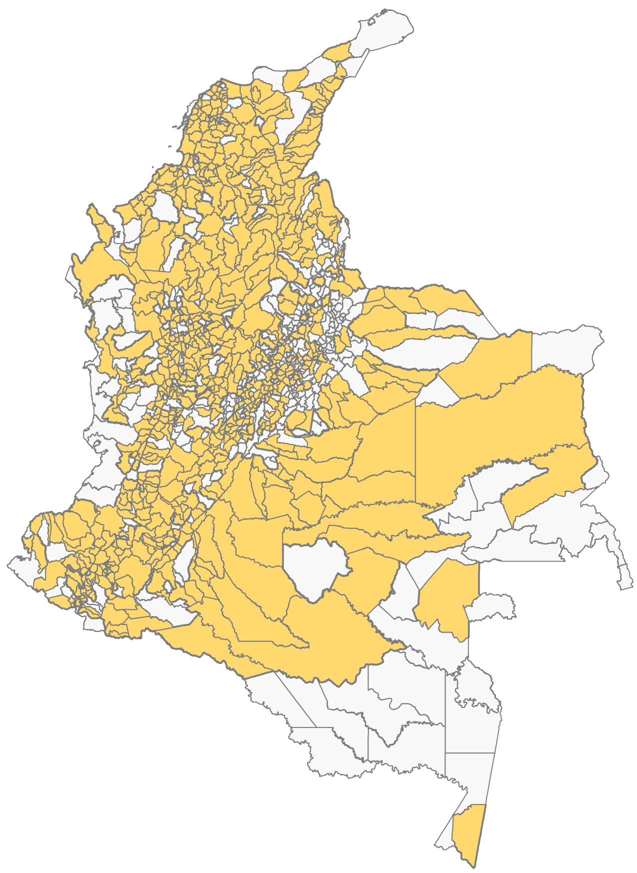
Procuraduría General de la Nación - Despacho Procurador General de la Nación
Carrera 5 No. 15-80 Bogotá, Colombia - PBX: 5878750 – procurador@procuraduria.gov.co
www.procuraduria.gov.co

Translation:

Respected Sir/Madam, Candidate to the Mayor's Office
Subject: *Campaign to promote citizens' oversight in the 2019 local elections*
The Attorney General of the Nation, in the exercise of its preventive functions, the Constitution, the Law and its Institutional Mission, is implementing a special program to watch over the forthcoming local elections of October 27 through an online campaign.
Making use of social media, a strategy to promote and strengthen citizens' use of an online reporting website, Pilas con el voto, administered by the Misión de Observación Electoral will be set in place. The goal of this strategy is to incentivize social control through citizen oversight and to guarantee transparency in the context of election day.
The Public Ministry welcomes your support, and thus we ask you to spread this information to your campaigns' offices and members. This same information will be communicated to the leaders of the other campaigns held in your municipality.[...]

Notes: On the left, we show the actual letter sent to candidates in the full knowledge condition. On the right is a translation to English of the text contained in the letter.

Figure A3: Municipalities in the Study Sample



Notes: This figure shows a map of Colombia with the administrative boundaries of municipalities. Municipalities in yellow were part of the study sample; those in light grey are not in the sample.

Table A1: Summary Statistics Comparing Study Sample to Average Municipality

	Study Sample			All Municipalities in Country		
	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	Min	Max	Mean	Minimum	Maximum
Population 2018 (Thousands)	27.22	7.20	139.36	37.16	0.41	2529.40
Facebook Penetration 2018	0.41	0.02	2.21	0.61	-	-
Per Capita GDP 2016 (Millions of Pesos)	13.92	2.55	349.12	13.72	2.49	349.12
% Rural Population 2017	52.36	1.65	97.99	55.88	0.09	100.00
% Poor 2005	45.39	6.84	100.00	45.03	5.36	100.00
Reports to MOE 2018	0.52	0.00	8.00	0.80	0.00	45.00
Reports to MOE 2015	3.43	0.00	33.00	3.83	0.00	152.00
Sample size		698			1121	

Notes: This table displays summary statistics for the municipalities in the study sample (columns 1-3) and the full set of municipalities in Colombia (columns 4-6) on a selected group of variables. Data for the average Facebook penetration rate for all of the country is averaged across the whole population, not across municipalities.

Table A2: Covariate Balance

	Control Mean	Any Treatment vs Control	Information vs Control	Call-to-action vs Control	Info + Call-to-action vs Control	Any Letter vs No Letter	Letter P. Knowledge vs No Letter	Letter F. Knowledge vs No Letter	Letter F. vs Letter P. Knowledge
Panel A. Previous Reports Covariates									
Reports to MOE 2018	0.502	0.097	0.138	0.017	0.134	-0.091	-0.177	-0.005	0.172
		(0.093)	(0.127)	(0.130)	(0.129)	(0.130)	(0.153)	(0.151)	(0.158)
Reports to MOE 2015	3.387	-0.159	0.308	-0.558	-0.235	0.248	0.063	0.433	0.370
		(0.433)	(0.577)	(0.513)	(0.532)	(0.456)	(0.512)	(0.581)	(0.606)
Panel B. Socioeconomic Covariates									
Log(Population)	9.959	-0.074	-0.049	-0.121	-0.054	-0.078	-0.074	-0.081	-0.006
		(0.071)	(0.086)	(0.088)	(0.091)	(0.078)	(0.091)	(0.089)	(0.091)
Facebook Penetration 2019	0.394	0.041	0.057	0.052	0.013	0.040	0.035	0.045	0.010
		(0.032)	(0.043)	(0.047)	(0.038)	(0.039)	(0.043)	(0.050)	(0.051)
GDP p,c. 2016 (Ms of Pesos)	14.513	-1.544	-2.917	-3.302	1.601	3.241	2.553	3.930	1.376
		(2.955)	(3.187)	(2.892)	(4.124)	(2.142)	(2.411)	(3.141)	(3.603)
Poor 2005 (%)	44.626	0.158	0.843	1.099	-1.473	-3.810	-4.490	-3.130	1.360
		(1.909)	(2.590)	(2.482)	(2.359)	(2.553)	(2.803)	(2.882)	(2.489)
Rural Population 2017	54.713	-1.733	-4.155	-1.288	0.283	0.554	-0.009	1.117	1.126
		(2.139)	(2.829)	(2.588)	(2.708)	(2.571)	(2.982)	(2.881)	(2.812)
Homicide Rate 2017	27.797	-0.529	-1.605	1.366	-1.322	1.722	-0.993	4.440	5.433
		(3.010)	(3.901)	(4.123)	(3.572)	(3.387)	(3.787)	(4.320)	(4.505)
Panel C. Political Covariates									
Number of Candidates 2019	5.916	-0.261	-0.243	-0.155	-0.384*	-0.278	-0.188	-0.368	-0.180
		(0.182)	(0.242)	(0.241)	(0.223)	(0.229)	(0.263)	(0.260)	(0.250)
Turnout 2015 (%)	64.185	0.138	1.041	0.026	-0.666	-0.426	0.077	-0.930	-1.007
		(0.816)	(1.061)	(1.023)	(1.091)	(0.958)	(1.125)	(1.158)	(1.244)
Share Blank Votes 2015 (%)	1.299	0.164	0.052	-0.069	0.510*	0.188	0.108	0.268	0.160

Continued on next page

Table A2 – continued from previous page

	Control Mean	Any Treatment vs Control	Information vs Control	Call-to-action vs Control	Info + Call-to-action vs Control	Any Letter vs No Letter	Letter P. Knowledge vs No Letter	Letter F. Knowledge vs No Letter	Letter F. vs Letter P. Knowledge
		(0.146)	(0.154)	(0.164)	(0.303)	(0.201)	(0.235)	(0.281)	(0.327)
Mayor Wining Margin 2015	13.482	-0.504	-0.566	-0.714	-0.231	0.947	0.562	1.333	0.771
		(1.057)	(1.446)	(1.253)	(1.370)	(1.274)	(1.468)	(1.483)	(1.485)
Santos Vote Share 2014	48.829	-0.813	1.599	-0.144	-3.927	-0.014	-1.017	0.990	2.007
		(2.078)	(2.619)	(2.602)	(2.544)	(2.194)	(2.676)	(2.569)	(2.877)
Zuluaga Vote Share 2014	48.548	0.623	-1.742	0.004	3.640	-0.087	0.999	-1.173	-2.172
		(2.028)	(2.548)	(2.538)	(2.487)	(2.133)	(2.612)	(2.488)	(2.802)
Liberals Vote Share 2018	15.520	-0.390	-0.999	-0.028	-0.132	0.159	-0.885	1.204	2.089
		(0.980)	(1.258)	(1.275)	(1.222)	(1.156)	(1.312)	(1.366)	(1.357)
Cambio R Vote Share 2018	15.259	-0.568	1.193	-1.242	-1.685	-0.714	-0.040	-1.388	-1.348
		(1.116)	(1.536)	(1.391)	(1.238)	(1.247)	(1.493)	(1.419)	(1.505)
Centro Dem Vote Share 2018	13.286	-0.680	-1.651	-1.112	0.735	0.704	2.155	-0.747	-2.902**
		(1.052)	(1.270)	(1.341)	(1.333)	(1.137)	(1.442)	(1.231)	(1.434)
P de la U Vote Share 2018	14.905	-0.226	-0.327	-0.103	-0.245	0.057	0.166	-0.052	-0.218
		(1.011)	(1.189)	(1.387)	(1.267)	(1.108)	(1.291)	(1.359)	(1.450)
Green Party Vote Share 2018	4.434	0.318	0.617	0.516	-0.183	-0.523	-0.780	-0.265	0.516
		(0.459)	(0.643)	(0.715)	(0.493)	(0.673)	(0.755)	(0.751)	(0.675)
Polo Vote Vote Share 2018	2.809	0.018	-0.158	0.565	-0.348	0.516	0.234	0.800	0.566
		(0.323)	(0.349)	(0.591)	(0.307)	(0.415)	(0.394)	(0.571)	(0.526)
Decentes Vote Share 2018	1.214	0.081	0.061	0.029	0.153	0.094	0.010	0.179	0.168
		(0.109)	(0.126)	(0.170)	(0.151)	(0.142)	(0.152)	(0.189)	(0.191)
Panel D. Geographic Covariates									
Caribean Region (=1)	0.204	-0.007	0.048	-0.009	-0.061	-0.002	0.010	-0.013	-0.022
		(0.038)	(0.052)	(0.049)	(0.045)	(0.044)	(0.053)	(0.050)	(0.053)
Center-East Region (=1)	0.222	-0.011	-0.018	0.001	-0.015	-0.000	-0.003	0.003	0.006
		(0.041)	(0.052)	(0.054)	(0.052)	(0.048)	(0.056)	(0.056)	(0.057)

Continued on next page

Table A2 – continued from previous page

	Control Mean	Any Treatment vs Control	Information vs Control	Call-to-action vs Control	Info + Call-to-action vs Control	Any Letter vs No Letter	Letter P. Knowledge vs No Letter	Letter F. Knowledge vs No Letter	Letter F. vs Letter P. Knowledge
Center-South Region (=1)	0.147	-0.027	-0.034	-0.063	0.015	-0.005	-0.009	-0.001	0.007
		(0.035)	(0.045)	(0.042)	(0.048)	(0.041)	(0.049)	(0.048)	(0.051)
Coffee-growing Region (=1)	0.160	0.075**	0.040	0.118**	0.068	0.021	0.013	0.028	0.016
		(0.033)	(0.043)	(0.047)	(0.045)	(0.045)	(0.053)	(0.053)	(0.055)
Llanos Region (=1)	0.058	0.007	-0.005	-0.018	0.045	0.018	0.024	0.011	-0.013
		(0.030)	(0.037)	(0.037)	(0.042)	(0.036)	(0.043)	(0.042)	(0.045)
Pacific Region (=1)	0.209	-0.038	-0.031	-0.030	-0.052	-0.031	-0.035	-0.028	0.007
		(0.034)	(0.043)	(0.042)	(0.041)	(0.040)	(0.044)	(0.045)	(0.041)
Panel E. Other Covariates									
Number Responses Survey	10.354	0.396	0.494	1.855**	-1.157*	0.330	0.370	0.289	-0.081
		(0.624)	(0.867)	(0.912)	(0.696)	(0.820)	(0.908)	(1.017)	(1.009)
Population Reached by Ads (%)	0.317	0.007	0.009	0.011	0.002	-0.002	0.001	-0.005	-0.007
		(0.015)	(0.020)	(0.020)	(0.020)	(0.019)	(0.023)	(0.022)	(0.024)
Users Reached by Ads (Thousands)	3.881	-0.707	-0.172	-1.053	-0.900	-0.491	-0.456	-0.525	-0.069
		(0.751)	(0.949)	(0.826)	(0.882)	(0.717)	(0.833)	(0.812)	(0.804)
No Users Reached by Ads (=1)	0.397	-0.002	-0.012	0.014	-0.006	-0.002	0.001	-0.006	-0.007
		(0.014)	(0.016)	(0.020)	(0.017)	(0.017)	(0.020)	(0.019)	(0.019)

Notes: This table presents the balance checks for a selected set of covariates. The control group mean of each variable is presented in column (1). In each of the remaining columns the difference in means is reported for the shown treatment groups. Observations are weighted by the percentage of the population older than 18 that was reached by a Facebook ad, except when reporting the difference in means of variables referring to the reach of the Facebook ads (i.e. the last three variables of Panel E). Robust standard errors are shown in parentheses; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A3: Balance on Pre-Treat Survey Respondent Characteristics

	Control Mean	Any Treatment vs Control	Information vs Control	Call-to-action vs Control	Info + Call-to-action vs Control	Any Letter vs No Letter	Letter P. Knowledge vs No Letter	Letter F. Knowledge vs No Letter	Letter F. vs Letter P. Knowledge
Female(=1)	0.507	0.008	-0.008	0.010	0.022	-0.014	-0.005	-0.022	-0.017
		(0.014)	(0.017)	(0.018)	(0.020)	(0.017)	(0.020)	(0.019)	(0.020)
Age	34.496	-0.206	-0.543	-0.185	0.142	0.979*	0.884	1.072*	0.188
		(0.379)	(0.531)	(0.494)	(0.477)	(0.500)	(0.567)	(0.585)	(0.571)
High School or Less (=1)	0.488	0.019	0.018	0.016	0.025	0.031*	0.032	0.030	-0.002
		(0.016)	(0.021)	(0.020)	(0.020)	(0.018)	(0.020)	(0.022)	(0.022)

Notes: This table presents the balance checks for a set of survey respondent characteristics. The control group mean of each variable is presented in column (1). In each of the remaining columns the difference in means is reported for the shown treatment groups. Clustered standard errors at the municipality level are shown in parentheses; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A4: Robustness of the Impacts on the Media-Based Irregularity Measures: Including News Coming From MOE Reports

	(1) Media Irregularities (=1)	(2) Number of Media Irregularities
Panel A. Pooled Treatment		
[T] Any treatment	-0.057*	-0.078*
	(0.032)	(0.045)
	[0.058]	[0.052]
Panel B. Subtreatments by Types of Ads		
[IA] Information Ad	-0.065	-0.083
	(0.041)	(0.061)
	[0.112]	[0.158]
[CA] Call-to-Action Ad	-0.021	-0.039
	(0.043)	(0.059)
	[0.632]	[0.488]
[$I+CA$] Info + Call-to-Action Ad	-0.083**	-0.113**
	(0.035)	(0.046)
	[0.018]	[0.038]
Test $IA = CA$, p-value	0.32	0.49
Test $IA = I + CA$, p-value	0.63	0.56
Test $CA = I + CA$, p-value	0.12	0.16
Panel C. Subtreatments by Letter - No Letter		
[NL] No Letter - Any Ad	-0.039	-0.067
	(0.041)	(0.055)
	[0.326]	[0.224]
[L] Letter - Any Ad	-0.066**	-0.084*
	(0.033)	(0.048)
	[0.042]	[0.084]
Test $NL = L$, p-value	0.47	0.72
Panel D. Subtreatments by Types of Letters		
[NL] No Letter - Any Ad	-0.039	-0.067
	(0.041)	(0.055)
	[0.334]	[0.276]
[PL] Partial Knowledge Letter - Any Ad	-0.052	-0.083
	(0.040)	(0.057)
	[0.208]	[0.178]
[FL] Full Knowledge Letter - Any Ad	-0.079**	-0.086
	(0.037)	(0.056)
	[0.038]	[0.118]
Test $NL = PL$, p-value	0.76	0.78
Test $NL = FL$, p-value	0.32	0.73
Test $PL = FL$, p-value	0.49	0.95
Control Mean	0.16	0.20
Sample Size	677	677

Notes: The outcome in column (1) is an indicator for whether any irregularity was reported in the media in a particular municipality. In column (2) it is the number of different irregularities. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A5: Impacts on Media-Based Irregularity Measures - By Type of Irregularity

	(1) Vote buying (=1)	(2) Riot (=1)	(3) Candidate intimidation (=1)	(4) Voter intimidation (=1)	(5) Registration fraud (=1)	(6) Public servant campaigning (=1)	(7) Electoral fraud (=1)	(8) Others (=1)
Panel A. Pooled Treatment								
[T] Any treatment	-0.045**	-0.029	-0.017	-0.013	0.004	0.002	-0.002	0.001
	(0.019)	(0.020)	(0.015)	(0.016)	(0.004)	(0.006)	(0.011)	(0.012)
	[0.000]	[0.062]	[0.212]	[0.338]	[0.638]	[0.792]	[0.848]	[0.908]
Panel B. Subtreatments by Types of Ads								
[IA] Information Ad	-0.049***	-0.029	-0.011	-0.010	0.001	-0.004	-0.001	0.005
	(0.019)	(0.022)	(0.020)	(0.019)	(0.002)	(0.004)	(0.016)	(0.017)
	[0.022]	[0.218]	[0.584]	[0.660]	[0.758]	[0.602]	[0.914]	[0.672]
[CA] Call to Action Ad	-0.034	-0.022	-0.012	-0.014	0.011	0.014	-0.005	0.002
	(0.022)	(0.023)	(0.020)	(0.018)	(0.012)	(0.014)	(0.012)	(0.016)
	[0.164]	[0.408]	[0.516]	[0.414]	[0.224]	[0.362]	[0.614]	[0.802]
[$I+CA$] Info + Call to Action Ad	-0.053***	-0.036*	-0.027*	-0.014	-0.000	-0.004	0.000	-0.004
	(0.018)	(0.020)	(0.014)	(0.017)	(0.002)	(0.004)	(0.012)	(0.012)
	[0.006]	[0.136]	[0.074]	[0.460]	[0.998]	[0.682]	[0.990]	[0.900]
Test $IA = CA$, p-value	0.26	0.74	0.96	0.79	0.38	0.16	0.77	0.84
Test $IA = I + CA$, p-value	0.32	0.61	0.29	0.76	0.60	.	0.94	0.57
Test $CA = I + CA$, p-value	0.13	0.43	0.32	0.99	0.35	0.16	0.63	0.73
Panel C. Subtreatments by Letter - No Letter								
[NL] No Letter - Any Ad	-0.047**	-0.027	-0.011	-0.014	0.012	0.003	0.003	0.012
	(0.019)	(0.022)	(0.020)	(0.017)	(0.012)	(0.009)	(0.016)	(0.017)
	[0.026]	[0.274]	[0.616]	[0.476]	[0.140]	[0.724]	[0.912]	[0.452]
[L] Letter - Any Ad	-0.044**	-0.030	-0.020	-0.012	-0.000	0.001	-0.005	-0.004
	(0.019)	(0.020)	(0.015)	(0.017)	(0.001)	(0.007)	(0.011)	(0.012)
	[0.006]	[0.108]	[0.174]	[0.512]	[0.844]	[0.908]	[0.680]	[0.692]
Test $NL = L$, p-value	0.77	0.79	0.57	0.87	0.26	0.81	0.61	0.31
Panel D. Subtreatments by Types of Letters								
[NL] No Letter - Any Ad	-0.047**	-0.027	-0.011	-0.014	0.012	0.003	0.003	0.012
	(0.019)	(0.022)	(0.020)	(0.017)	(0.012)	(0.009)	(0.016)	(0.018)
	[0.044]	[0.254]	[0.632]	[0.478]	[0.122]	[0.698]	[0.878]	[0.494]
[PL] Partial Knowledge Letter - Any Ad	-0.053***	-0.037*	-0.020	0.001	-0.001	-0.004	-0.000	0.003
	(0.018)	(0.021)	(0.018)	(0.022)	(0.001)	(0.004)	(0.014)	(0.016)
	[0.008]	[0.144]	[0.318]	[0.954]	[0.576]	[0.616]	[0.996]	[0.718]
[FL] Full Knowledge Letter - Any Ad	-0.036*	-0.024	-0.020	-0.026*	0.000	0.006	-0.009	-0.012
	(0.022)	(0.022)	(0.016)	(0.015)	(0.002)	(0.012)	(0.011)	(0.010)
	[0.128]	[0.306]	[0.248]	[0.110]	[0.942]	[0.700]	[0.486]	[0.320]
Test $NL = PL$, p-value	0.32	0.52	0.63	0.40	0.23	0.32	0.86	0.66
Test $NL = FL$, p-value	0.41	0.87	0.57	0.14	0.30	0.81	0.40	0.10
Test $PL = FL$, p-value	0.15	0.45	0.97	0.10	0.06	0.32	0.41	0.24
Control Mean	0.05	0.04	0.03	0.01	0.01	0.00	0.01	0.01
Sample Size	677	677	677	677	677	677	677	677

Notes: The outcomes in columns (1)-(8) are indicators for whether each of the types of irregularities displayed were reported in the media. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A6: Robustness of the Media-Based Irregularity Measures: Leave-one-out Estimates

	(1) Any irregularity except vote buying (=1)	(2) Any irregularity except riots (=1)	(3) Any irregularity except candidate intimidation (=1)	(4) Any irregularity except voter intimidation (=1)	(5) Any irregularity except registration fraud (=1)	(6) Any irregularity except public servant campaigning (=1)	(7) Any irregularity except electoral fraud (=1)	(8) Any irregularity except others (=1)
Panel A. Pooled Treatment								
[T] Any treatment	-0.048	-0.064**	-0.060**	-0.076**	-0.087***	-0.084***	-0.077**	-0.078**
	(0.030)	(0.029)	(0.030)	(0.031)	(0.032)	(0.032)	(0.031)	(0.031)
	[0.084]	[0.020]	[0.014]	[0.004]	[0.002]	[0.000]	[0.004]	[0.004]
Panel B. Subtreatments by Types of Ads								
[IA] Information Ad	-0.035	-0.055	-0.057	-0.069*	-0.071*	-0.069*	-0.068*	-0.074**
	(0.038)	(0.037)	(0.036)	(0.037)	(0.040)	(0.040)	(0.038)	(0.037)
	[0.344]	[0.140]	[0.140]	[0.096]	[0.112]	[0.104]	[0.064]	[0.062]
[CA] Call-to-Action Ad	-0.028	-0.047	-0.032	-0.053	-0.070*	-0.065	-0.052	-0.046
	(0.039)	(0.037)	(0.039)	(0.039)	(0.040)	(0.041)	(0.040)	(0.040)
	[0.518]	[0.252]	[0.430]	[0.202]	[0.094]	[0.120]	[0.222]	[0.244]
[$I+CA$] Info + Call-to-Action Ad	-0.080***	-0.089***	-0.091***	-0.106***	-0.119***	-0.117***	-0.110***	-0.114***
	(0.030)	(0.029)	(0.030)	(0.030)	(0.032)	(0.032)	(0.031)	(0.031)
	[0.010]	[0.008]	[0.012]	[0.002]	[0.000]	[0.000]	[0.000]	[0.002]
Test $IA = CA$, p-value	0.85	0.81	0.50	0.67	0.97	0.92	0.68	0.46
Test $IA = I + CA$, p-value	0.14	0.24	0.22	0.18	0.12	0.12	0.14	0.14
Test $CA = I + CA$, p-value	0.10	0.15	0.06	0.08	0.11	0.10	0.07	0.03
Panel C. Subtreatments by Letter - No Letter								
[NL] No Letter - Any Ad	-0.029	-0.044	-0.038	-0.050	-0.074*	-0.069*	-0.060	-0.060
	(0.037)	(0.037)	(0.038)	(0.038)	(0.038)	(0.039)	(0.038)	(0.038)
	[0.486]	[0.228]	[0.312]	[0.196]	[0.052]	[0.108]	[0.154]	[0.128]
[L] Letter - Any Ad	-0.057*	-0.074**	-0.072**	-0.089***	-0.093***	-0.092***	-0.085***	-0.088***
	(0.031)	(0.030)	(0.031)	(0.031)	(0.033)	(0.033)	(0.032)	(0.032)
	[0.048]	[0.010]	[0.014]	[0.004]	[0.006]	[0.006]	[0.002]	[0.002]
Test $NL = L$, p-value	0.36	0.31	0.25	0.20	0.54	0.45	0.38	0.37
Panel D. Subtreatments by Types of Letters								
[NL] No Letter - Any Ad	-0.029	-0.044	-0.038	-0.050	-0.074*	-0.069*	-0.060	-0.060
	(0.037)	(0.037)	(0.038)	(0.038)	(0.039)	(0.039)	(0.038)	(0.038)
	[0.432]	[0.286]	[0.354]	[0.202]	[0.056]	[0.088]	[0.126]	[0.150]
[PL] Partial Knowledge Letter - Any Ad	-0.043	-0.058	-0.061*	-0.091**	-0.081**	-0.081**	-0.080**	-0.084**
	(0.038)	(0.036)	(0.036)	(0.036)	(0.039)	(0.039)	(0.038)	(0.037)
	[0.308]	[0.124]	[0.126]	[0.008]	[0.048]	[0.048]	[0.042]	[0.032]
[FL] Full Knowledge Letter - Any Ad	-0.072**	-0.089***	-0.084***	-0.088***	-0.105***	-0.104***	-0.090***	-0.092***
	(0.032)	(0.030)	(0.032)	(0.033)	(0.035)	(0.035)	(0.034)	(0.034)
	[0.040]	[0.006]	[0.030]	[0.020]	[0.012]	[0.010]	[0.016]	[0.016]
Test $NL = PL$, p-value	0.72	0.69	0.52	0.24	0.87	0.75	0.56	0.51
Test $NL = FL$, p-value	0.18	0.12	0.14	0.25	0.33	0.28	0.33	0.33
Test $PL = FL$, p-value	0.36	0.28	0.45	0.90	0.45	0.48	0.75	0.79
Control Mean	0.11	0.12	0.11	0.13	0.14	0.14	0.13	0.13
Sample Size	677	677	677	677	677	677	677	677

Notes: The outcome in each column is an indicator for whether any irregularity was reported in the media in a particular municipality, when one leaves out each of the types of irregularities shown in columns (1)-(8). All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A7: Correlation Between Forensic and Media-Based Electoral Irregularity Measures

	(1)	(2)	(3)	(4)	(5)	(6)
	Index of all Forensic Test Stats (z-score)		Any P-value From Forensic Tests < 0.1 (=1)		Any P-value From Forensic Tests < 0.05 (=1)	
Media Irregularities (=1)	0.148*		0.079		0.142**	
	(0.079)		(0.058)		(0.058)	
Number of Media Irregularities		0.187*		0.101		0.180**
		(0.100)		(0.073)		(0.073)
Sample Size	218	218	218	218	218	218

Notes: This table presents the OLS results of regressing z-scores of the forensic variables detailed in Table 4 on z-scores of the media-based measure of irregularities detailed in Table 3. The sample is restricted to municipalities in the control group. Since all of these variables are normalized so the estimates reported can be interpreted as correlations. Robust standard errors are shown in parentheses; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A8: Impacts on Deviations from Benford's Second Digit Law - By Test

Test:	(1)	(2) Pearson χ^2	(3)	(4)	(5) Kolmogorov-Smirnov D	(6)	(7)	(8) Kuiper V	(9)
	Stat (z-score)	P-value < 0.1 (=1)	P-value < 0.05 (=1)	Stat (z-score)	P-value < 0.1 (=1)	P-value < 0.05 (=1)	Stat (z-score)	P-value < 0.1 (=1)	P-value < 0.05 (=1)
Panel A. Pooled Treatment									
[T] Any treatment	-0.052	-0.059	-0.028	-0.192**	-0.109**	-0.133***	-0.215***	-0.095**	-0.086**
	(0.089)	(0.046)	(0.044)	(0.080)	(0.046)	(0.045)	(0.075)	(0.046)	(0.044)
	[0.502]	[0.202]	[0.514]	[0.012]	[0.018]	[0.002]	[0.002]	[0.034]	[0.046]
Panel B. Subtreatments by Types of Ads									
[IA] Information Ad	-0.090	-0.079	-0.043	-0.339***	-0.151***	-0.165***	-0.307***	-0.166***	-0.122**
	(0.114)	(0.058)	(0.056)	(0.102)	(0.058)	(0.057)	(0.093)	(0.056)	(0.054)
	[0.438]	[0.186]	[0.460]	[0.004]	[0.006]	[0.000]	[0.002]	[0.002]	[0.036]
[CA] Call-to-Action Ad	-0.014	-0.027	-0.022	-0.095	-0.073	-0.093	-0.144	-0.025	-0.038
	(0.107)	(0.059)	(0.057)	(0.101)	(0.059)	(0.058)	(0.093)	(0.059)	(0.056)
	[0.904]	[0.620]	[0.728]	[0.344]	[0.228]	[0.114]	[0.128]	[0.668]	[0.496]
[$I+CA$] Info + Call-to-Action Ad	-0.051	-0.071	-0.019	-0.138	-0.102*	-0.141**	-0.192**	-0.092	-0.099*
	(0.108)	(0.058)	(0.055)	(0.102)	(0.059)	(0.057)	(0.095)	(0.058)	(0.053)
	[0.642]	[0.250]	[0.758]	[0.170]	[0.076]	[0.012]	[0.044]	[0.112]	[0.058]
Test $IA = CA$, p-value	0.51	0.42	0.72	0.03	0.21	0.25	0.09	0.02	0.14
Test $IA = I + CA$, p-value	0.73	0.91	0.69	0.06	0.44	0.70	0.23	0.22	0.68
Test $CA = I + CA$, p-value	0.73	0.49	0.97	0.69	0.65	0.44	0.62	0.28	0.28
Panel C. Subtreatments by Letter - No Letter									
[NL] No Letter - Any Ad	-0.073	-0.067	-0.026	-0.112	-0.077	-0.090	-0.208**	-0.085	-0.107**
	(0.106)	(0.058)	(0.055)	(0.099)	(0.059)	(0.058)	(0.090)	(0.058)	(0.054)
	[0.532]	[0.270]	[0.648]	[0.288]	[0.228]	[0.144]	[0.026]	[0.138]	[0.062]
[L] Letter - Any Ad	-0.041	-0.056	-0.029	-0.234***	-0.125**	-0.155***	-0.220***	-0.100**	-0.076
	(0.096)	(0.049)	(0.047)	(0.088)	(0.049)	(0.048)	(0.082)	(0.049)	(0.047)
	[0.708]	[0.264]	[0.536]	[0.014]	[0.018]	[0.000]	[0.012]	[0.034]	[0.088]
Test $NL = L$, p-value	0.73	0.83	0.94	0.20	0.39	0.23	0.88	0.78	0.52
Panel D. Subtreatments by Types of Letters									
[NL] No Letter - Any Ad	-0.073	-0.067	-0.026	-0.111	-0.077	-0.090	-0.208**	-0.085	-0.107**
	(0.106)	(0.058)	(0.055)	(0.099)	(0.059)	(0.058)	(0.090)	(0.058)	(0.054)
	[0.564]	[0.216]	[0.668]	[0.258]	[0.200]	[0.136]	[0.042]	[0.160]	[0.064]
[PL] Partial Knowledge Letter - Any Ad	-0.053	-0.060	-0.029	-0.334***	-0.173***	-0.197***	-0.255***	-0.112*	-0.091*
	(0.112)	(0.060)	(0.057)	(0.090)	(0.058)	(0.056)	(0.088)	(0.058)	(0.054)
	[0.648]	[0.344]	[0.616]	[0.002]	[0.000]	[0.002]	[0.008]	[0.064]	[0.114]
[FL] Full Knowledge Letter - Any Ad	-0.029	-0.051	-0.030	-0.131	-0.077	-0.114**	-0.184*	-0.088	-0.061
	(0.114)	(0.058)	(0.055)	(0.114)	(0.058)	(0.057)	(0.102)	(0.056)	(0.055)
	[0.806]	[0.398]	[0.582]	[0.240]	[0.178]	[0.036]	[0.068]	[0.140]	[0.266]
Test $NL = PL$, p-value	0.85	0.92	0.96	0.02	0.13	0.08	0.59	0.67	0.77
Test $NL = FL$, p-value	0.68	0.79	0.95	0.86	1.00	0.71	0.82	0.96	0.41
Test $PL = FL$, p-value	0.83	0.88	0.99	0.06	0.12	0.17	0.47	0.70	0.60
Control Mean	0.00	0.38	0.28	0.00	0.48	0.42	0.00	0.45	0.34
Sample Size	677	677	677	677	677	677	677	677	677

Notes: This table reports the effects of the intervention on the χ^2, Kolmogorov-Smirnov and Kuiper test statistics testing for Benford's 2nd digit law. For each test, the outcome is the test statistic, as well as indicators that take the value of one if the p-value of each test leads to rejection of the null hypothesis with less than a 10% or 5% significance level. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A9: Covariate Balance - Candidate Level Data

	Control Mean	Any Treatment vs Control	Information vs Control	Call-to-action vs Control	Info + Call-to-action vs Control	Any Letter vs No Letter	Letter P. Knowledge vs No Letter	Letter F. Knowledge vs No Letter	Letter F. vs Letter P. Knowledge
Panel A. Previous Reports Covariates									
Reports to MOE 2018	0.534	0.143	0.228	0.001	0.205	-0.050	-0.112	0.014	0.125
		(0.111)	(0.162)	(0.148)	(0.151)	(0.154)	(0.192)	(0.172)	(0.196)
Reports to MOE 2015	3.541	0.173	0.745	-0.282	0.056	0.354	0.212	0.500	0.288
		(0.510)	(0.709)	(0.629)	(0.670)	(0.607)	(0.722)	(0.738)	(0.810)
Panel B. Socioeconomic Covariates									
Log(Population)	10.107	-0.087	-0.071	-0.132	-0.056	-0.053	-0.033	-0.073	-0.040
		(0.081)	(0.101)	(0.100)	(0.102)	(0.090)	(0.106)	(0.103)	(0.107)
Facebook Penetration 2019	0.448	0.038	0.053	0.052	0.007	0.020	0.024	0.015	-0.009
		(0.035)	(0.047)	(0.047)	(0.040)	(0.042)	(0.048)	(0.050)	(0.051)
GDP p,c. 2016 (Ms of Pesos)	15.236	-1.572	-2.808	-2.786	1.023	2.398	2.201	2.598	0.397
		(2.854)	(3.077)	(2.899)	(3.690)	(2.004)	(2.390)	(2.563)	(2.908)
Poor 2005 (%)	43.006	-0.138	0.468	1.024	-2.014	-2.744	-4.456	-0.993	3.462
		(2.080)	(2.845)	(2.684)	(2.511)	(2.722)	(2.996)	(3.086)	(2.727)
Rural Population 2017	49.379	-0.030	-2.425	0.193	2.258	2.297	0.688	3.943	3.255
		(2.428)	(3.195)	(2.997)	(3.012)	(2.877)	(3.307)	(3.245)	(3.159)
Homicide Rate 2017	28.994	-0.798	-0.642	-0.192	-1.608	1.520	-0.165	3.243	3.407
		(2.966)	(3.949)	(3.805)	(3.469)	(3.378)	(3.769)	(4.126)	(4.091)
Panel C. Political Covariates									
Number of Candidates 2019	6.744	-0.206	-0.169	-0.074	-0.387	-0.413	-0.349	-0.479	-0.129
		(0.231)	(0.319)	(0.332)	(0.269)	(0.317)	(0.365)	(0.356)	(0.343)
Turnout 2015 (%)	63.028	0.255	1.102	0.378	-0.768	-0.668	-0.321	-1.023	-0.702
		(0.874)	(1.168)	(1.157)	(1.114)	(1.086)	(1.232)	(1.330)	(1.357)
Share Blank Votes 2015 (%)	1.402	0.236	0.104	-0.066	0.697	0.240	0.132	0.350	0.219

Continued on next page

Table A9 – continued from previous page

	Control Mean	Any Treatment vs Control	Information vs Control	Call-to-action vs Control	Info + Call-to-action vs Control	Any Letter vs No Letter	Letter P. Knowledge vs No Letter	Letter F. Knowledge vs No Letter	Letter F. vs Letter P. Knowledge
		(0.189)	(0.194)	(0.193)	(0.434)	(0.272)	(0.256)	(0.436)	(0.461)
Mayor Wining Margin 2015	13.455	-0.318	-0.461	-0.324	-0.161	0.360	0.547	0.169	-0.378
		(1.146)	(1.619)	(1.392)	(1.397)	(1.413)	(1.628)	(1.606)	(1.572)
Santos Vote Share 2014	48.549	-1.557	0.002	-0.608	-4.210	1.070	-1.318	3.513	4.830*
		(2.112)	(2.641)	(2.683)	(2.637)	(2.296)	(2.727)	(2.690)	(2.917)
Zuluaga Vote Share 2014	48.734	1.309	-0.246	0.490	3.821	-1.075	1.319	-3.525	-4.844*
		(2.056)	(2.552)	(2.609)	(2.581)	(2.215)	(2.655)	(2.585)	(2.843)
Liberals Vote Share 2018	15.064	-0.899	-1.628	-0.612	-0.437	0.658	-0.476	1.818	2.294
		(0.996)	(1.194)	(1.283)	(1.338)	(1.144)	(1.294)	(1.395)	(1.419)
Cambio R Vote Share 2018	15.696	-1.165	0.362	-1.610	-2.300	-0.658	-0.664	-0.652	0.011
		(1.314)	(1.703)	(1.674)	(1.426)	(1.401)	(1.572)	(1.679)	(1.648)
Centro Dem Vote Share 2018	13.710	0.029	-0.536	-0.571	1.263	-0.221	1.786	-2.274	-4.060***
		(1.130)	(1.586)	(1.446)	(1.408)	(1.512)	(1.760)	(1.604)	(1.519)
P de la U Vote Share 2018	14.859	-0.156	-0.269	0.875	-1.135	0.502	0.329	0.679	0.349
		(1.016)	(1.191)	(1.487)	(1.235)	(1.169)	(1.338)	(1.487)	(1.587)
Green Party Vote Share 2018	4.365	0.341	0.743	0.113	0.160	-0.677	-0.886	-0.463	0.422
		(0.498)	(0.701)	(0.679)	(0.594)	(0.686)	(0.755)	(0.787)	(0.703)
Polo Vote Vote Share 2018	2.988	-0.085	-0.234	0.327	-0.367	0.780**	0.473	1.094**	0.621
		(0.353)	(0.390)	(0.552)	(0.351)	(0.334)	(0.320)	(0.519)	(0.544)
Decentes Vote Share 2018	1.289	0.123	0.068	0.107	0.197	0.136	0.035	0.238	0.203
		(0.124)	(0.141)	(0.210)	(0.164)	(0.163)	(0.167)	(0.232)	(0.239)
Panel D. Geographic Covariates									
Caribean Region (=1)	0.227	-0.026	0.022	-0.000	-0.104**	0.011	-0.005	0.028	0.033
		(0.044)	(0.057)	(0.058)	(0.047)	(0.048)	(0.054)	(0.059)	(0.059)
Center-East Region (=1)	0.222	-0.005	-0.030	-0.009	0.026	0.011	0.013	0.010	-0.004
		(0.044)	(0.054)	(0.057)	(0.060)	(0.053)	(0.062)	(0.060)	(0.061)

Continued on next page

Table A9 – continued from previous page

	Control Mean	Any Treatment vs Control	Information vs Control	Call-to-action vs Control	Info + Call-to-action vs Control	Any Letter vs No Letter	Letter P. Knowledge vs No Letter	Letter F. Knowledge vs No Letter	Letter F. vs Letter P. Knowledge
Center-South Region (=1)	0.169	-0.024	-0.031	-0.067	0.029	-0.019	-0.018	-0.019	-0.002
		(0.037)	(0.047)	(0.045)	(0.054)	(0.047)	(0.054)	(0.055)	(0.053)
Coffee-growing Region (=1)	0.153	0.084**	0.053	0.129**	0.068	0.041	0.051	0.030	-0.020
		(0.037)	(0.047)	(0.054)	(0.049)	(0.049)	(0.060)	(0.056)	(0.063)
Llanos Region (=1)	0.071	0.013	0.015	-0.009	0.036	-0.001	0.016	-0.018	-0.034
		(0.034)	(0.049)	(0.043)	(0.044)	(0.046)	(0.055)	(0.050)	(0.049)
Pacific Region (=1)	0.158	-0.042	-0.028	-0.043	-0.055	-0.044	-0.057	-0.031	0.027
		(0.035)	(0.045)	(0.042)	(0.041)	(0.039)	(0.042)	(0.045)	(0.038)
Panel E. Other Covariates									
Number Responses Survey	10.945	0.384	0.280	1.943**	-1.167	-0.161	0.081	-0.409	-0.490
		(0.642)	(0.850)	(0.984)	(0.729)	(0.869)	(0.923)	(1.111)	(1.067)
Population Reached by Ads (%)	0.353	0.011	0.009	0.020	0.003	-0.001	0.007	-0.009	-0.016
		(0.016)	(0.020)	(0.021)	(0.020)	(0.020)	(0.023)	(0.023)	(0.024)
Users Reached by Ads (Thousands)	8.126	-0.950	-0.239	-1.377	-1.254	-0.506	-0.158	-0.846	-0.689
		(1.054)	(1.376)	(1.132)	(1.229)	(1.009)	(1.234)	(1.101)	(1.179)
No Users Reached by Ads (=1)	0.001	0.002	0.000***	0.003	0.002	0.000	-0.001	0.002	0.003
		(0.001)	(0.000)	(0.003)	(0.002)	(0.002)	(0.001)	(0.003)	(0.003)
Panel F. Candidate Level Covariates									
Candidate Will Engage in Irregularities (%)	0.511	0.009	0.023	-0.017	0.020	-0.010	-0.000	-0.020	-0.020
		(0.018)	(0.022)	(0.023)	(0.023)	(0.021)	(0.024)	(0.024)	(0.024)
Demeaned - Candidate Will Engage in Irregularities (%)	-0.000	0.000	0.000	0.000*	0.000	0.000	0.000	0.000	-0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Above Average - Candidate Will Engage in Irregularities (%)	0.469	0.010	0.014	0.010	0.005	0.014	-0.003	0.031	0.034*
		(0.016)	(0.021)	(0.020)	(0.019)	(0.017)	(0.020)	(0.019)	(0.021)

Continued on next page

Table A9 – continued from previous page

	Control Mean	Any Treatment vs Control	Information vs Control	Call-to-action vs Control	Info + Call-to-action vs Control	Any Letter vs No Letter	Letter P. Knowledge vs No Letter	Letter F. Knowledge vs No Letter	Letter F. vs Letter P. Knowledge
Past Malfeasance (=1)	0.263	-0.014	0.009	-0.047	-0.011	-0.025	-0.066	0.052	0.118*
		(0.051)	(0.065)	(0.070)	(0.045)	(0.065)	(0.068)	(0.062)	(0.061)
Log(Age)	3.791	0.010	0.004	0.013	0.012	-0.000	0.009	-0.009	-0.018
		(0.009)	(0.013)	(0.011)	(0.011)	(0.011)	(0.013)	(0.013)	(0.013)
Female(=1)	0.172	-0.031**	-0.052***	-0.003	-0.038**	-0.010	-0.010	-0.011	-0.000
		(0.015)	(0.018)	(0.020)	(0.017)	(0.016)	(0.018)	(0.020)	(0.021)
Incumbent Party - Lax (=1)	0.184	-0.002	-0.007	-0.002	0.003	0.025*	0.031*	0.019	-0.012
		(0.013)	(0.018)	(0.014)	(0.016)	(0.014)	(0.016)	(0.017)	(0.016)
Incumbent Party - Strict (=1)	0.044	0.010	-0.005	0.012	0.024**	0.003	-0.001	0.008	0.009
		(0.008)	(0.010)	(0.011)	(0.011)	(0.010)	(0.012)	(0.012)	(0.013)
Party Coalition (=1)	0.245	-0.002	-0.015	-0.000	0.010	0.012	0.010	0.013	0.003
		(0.021)	(0.030)	(0.026)	(0.026)	(0.025)	(0.029)	(0.029)	(0.030)
Independent Candidate (=1)	0.027	-0.010	-0.010	0.003	-0.024***	0.008	0.014	0.001	-0.012
		(0.007)	(0.008)	(0.010)	(0.007)	(0.007)	(0.009)	(0.008)	(0.009)

Notes: This table presents the balance checks for a selected set of covariates in the candidate-level data. The control group mean of each variable is presented in column (1). In each of the remaining columns the difference in means is reported for the shown treatment groups. Observations are weighted by the percentage of the population older than 18 that was reached by a Facebook ad, except when reporting the difference in means of variables referring to the reach of the Facebook ads (i.e. the last three variables of Panel E). Clustered standard errors at the municipality level are shown in parentheses; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A10: Correlation Between Past Malfeasance and Survey Measures of Likelihood to Engage in Irregularities

	(1)	(2)	(3)
	Past Malfeasance (z-score)		
Candidate will engage in irregularities (z-score)	0.272***		
	(0.057)		
Demeaned Candidate will engage in irregularities (z-score)		0.412***	
		(0.056)	
Above Average Candidate will engage in irregularities (z-score)			0.383***
			(0.060)
Sample Size	256	256	256
N. Municipalities	48	48	48

Notes: This table presents the OLS results of regressing a z-score of an indicator for whether a candidate was found to be involved in malfeasance in the past according to the investigation by the NGO PARES on different survey-based variables about the likelihood that a candidate commits irregularities. All of these variables are normalized so the estimates reported can be interpreted as correlations. In column (1) the dependent variable is the proportion of respondents from the pre-treatment survey that say that the candidate might commit at least one type of electoral irregularity. In column (2) it is this same variable, demeaned using the municipality-level mean. In column (3) it is an indicator that takes the value of one if this variable is above the municipal-level mean. Robust standard errors are shown in parentheses; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A11: Correlation Between Measures of Candidate Engagement in Irregularities and Other Covariates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Candidate Will Win the Election (fraction of respondents)			Incumbent Party Candidate Strict Measure (=1)			Incumbent Party Candidate Lax Measure (=1)		
Candidate will engage in irregularities (z-score)	0.088***			0.102***			0.174***		
	(0.019)			(0.018)			(0.018)		
Demeaned Candidate will engage in irregularities (z-score)		0.114***			0.134***			0.208***	
		(0.019)			(0.022)			(0.018)	
Above Average Candidate will engage in irregularities (z-score)			0.118***			0.122***			0.182***
			(0.018)			(0.019)			(0.018)
Sample Size	2986	2986	2986	2986	2986	2986	2986	2986	2986
N. Municipalities	628	628	628	628	628	628	628	628	628

Notes: This table presents the OLS results of regressing the z-scores of the measures indicating each candidates' likelihood of engagement in electoral irregularities from Table **??** on z-scores of different candidate covariates. In columns (1)-(3) the examined covariate is the proportion of respondents from the pre-treatment survey that say that the candidate is going to win the election in their municipality. In columns (4)-(6) it is an indicator that takes the value of one if the candidate belongs to exactly the same party or coalition of parties as the incumbent mayor. In columns (7)-(9) it is an indicator that takes the value of one if the candidate belongs to a party or coalition of parties that share at least one party with the incumbent mayor. Since all of these variables are normalized so the estimates reported can be interpreted as correlations. Robust standard errors are shown in parentheses; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A12: Impacts on Vote Share of More Popular Candidates

	(1)	(2)	(3)
		Vote Share (%)	
Interaction term Z:	Candidate Will Win the Election (fraction of respondents)	Incumbent Party Candidate Strict Measure (=1)	Incumbent Party Candidate Lax Measure (=1)
Panel A. Pooled Treatment			
[$T \times Z$] Any treatment $\times Z$	-1.557	-1.750	1.804
	(1.818)	(2.862)	(1.696)
	[0.330]	[0.518]	[0.280]
Panel B. Subtreatments by Types of Ads			
[$IA \times Z$] Information Ad $\times Z$	-2.198	-6.083	2.080
	(2.637)	(5.030)	(2.339)
	[0.390]	[0.106]	[0.376]
[$CA \times Z$] Call-to-Action Ad $\times Z$	0.619	-3.973	1.336
	(2.179)	(3.441)	(2.073)
	[0.804]	[0.228]	[0.520]
[$I + CA \times Z$] Info + Call-to-Action Ad $\times Z$	-3.135	3.015	2.024
	(2.308)	(2.960)	(2.017)
	[0.172]	[0.332]	[0.324]
Test $IA \times Z = CA \times Z$, p-value	0.30	0.68	0.76
Test $IA \times Z = I + CA \times Z$, p-value	0.74	0.06	0.98
Test $CA \times Z = I + CA \times Z$, p-value	0.12	0.02	0.74
Panel C. Subtreatments by Letter - No Letter			
[$NL \times Z$] No Letter - Any Ad $\times Z$	-1.292	-2.933	0.230
	(2.618)	(3.555)	(2.013)
	[0.654]	[0.418]	[0.896]
[$L \times Z$] Letter - Any Ad $\times Z$	-1.696	-1.138	2.583
	(1.893)	(3.128)	(1.853)
	[0.360]	[0.746]	[0.178]
Test $NL \times Z = L \times Z$, p-value	0.87	0.60	0.21
Panel D. Subtreatments by Types of Letters			
[$NL \times Z$] No Letter - Any Ad $\times Z$	-1.292	-2.938	0.227
	(2.619)	(3.556)	(2.014)
	[0.622]	[0.472]	[0.912]
[$FL \times Z$] Full Knowledge Letter - Any Ad $\times Z$	0.112	0.185	4.931**
	(2.330)	(3.651)	(2.273)
	[0.956]	[0.956]	[0.028]
[$PL \times Z$] Partial Knowledge Letter - Any Ad $\times Z$	-3.516	-2.690	0.406
	(2.149)	(3.797)	(2.081)
	[0.102]	[0.478]	[0.832]
Test $NL \times Z = PL \times Z$, p-value	0.41	0.95	0.93
Test $NL \times Z = FL \times Z$, p-value	0.62	0.42	0.04
Test $PL \times Z = FL \times Z$, p-value	0.13	0.48	0.05
Control Mean	19.05	19.05	19.05
Sample Size	2986	2986	2986
N. Municipalities	628	628	628

Notes: The outcome in all columns is the vote share for each candidate, expressed as a percentage of total valid votes. In each of these columns, different proxies for candidate popularity are used to compute candidate-level heterogenous effects. In column (1) it is the proportion of respondents from the pre-treatment survey that say that the candidate is going to win the election in their municipality. In column (2) it is an indicator that takes the value of one if the candidate belongs to exactly the same party or coalition of parties as the incumbent mayor. In column (3) it is an indicator that takes the value of one if the candidate belongs to a party or coalition of parties that share at least one party with the incumbent mayor. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Clustered standard errors at the municipality level are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A13: Estimates of the Percentage of the Effect on Candidate Vote Share Due to Decreasing Electoral Irregularities

Electoral Irregularity Measure / Measure of Z_c:	Candidate will engage in irregularities (fraction of respondents)	Demeaned Candidate will engage in irregularities (fraction of respondents)	Above Average Candidate will engage in irregularities (=1)
Media Irregularities (=1)	13.06	15.63	15.58
Number of Media Irregularities	13.23	11.83	13.36
Index of all Forensic Test Stats (z-score)	35.22	20.39	25.29
Any P-value From Forensic Tests < 0.1 (=1)	19.80	22.77	20.30
Any P-value From Forensic Tests < 0.05 (=1)	32.96	31.94	26.56

Notes: This table presents estimates of the percentage of the effect of the reporting campaign interventions on the vote share of the candidates more likely to engage in electoral irregularities accounted by the decrease in electoral irregularities using the method described in Section 4.4, using different combinations of variables proxing for electoral irregularities and for the proxies for the likelihood that a candidate engages in electoral irregularities, Z_c.

Table A14: Robustness: Impacts on Reports - No Controls

	(1) Reports (=1)	(2) N. Reports	(3) High Quality Reports (=1)	(4) High Quality N. Reports
Panel A. Pooled Treatment				
$[T]$ Any treatment	0.106**	0.373***	0.080**	0.186***
	(0.045)	(0.135)	(0.038)	(0.067)
	[0.022]	[0.016]	[0.064]	[0.012]
Panel B. Subtreatments by Types of Ads				
$[IA]$ Information Ad	0.159***	0.517***	0.066	0.122
	(0.058)	(0.189)	(0.049)	(0.075)
	[0.002]	[0.002]	[0.150]	[0.088]
$[CA]$ Call-to-Action Ad	-0.024	0.087	0.055	0.144
	(0.057)	(0.171)	(0.049)	(0.096)
	[0.694]	[0.616]	[0.252]	[0.110]
$[I+CA]$ Info + Call-to-Action Ad	0.182***	0.511***	0.117**	0.292***
	(0.059)	(0.190)	(0.052)	(0.110)
	[0.006]	[0.006]	[0.024]	[0.006]
Test $IA = CA$, p-value	0.00	0.04	0.83	0.83
Test $IA = I + CA$, p-value	0.73	0.98	0.37	0.15
Test $CA = I + CA$, p-value	0.00	0.05	0.28	0.26
Panel C. Subtreatments by Letter - No Letter				
$[NL]$ No Letter - Any Ad	0.217***	0.484***	0.115**	0.212**
	(0.059)	(0.171)	(0.051)	(0.099)
	[0.000]	[0.000]	[0.026]	[0.012]
$[L]$ Letter - Any Ad	0.049	0.315**	0.061	0.173**
	(0.048)	(0.155)	(0.041)	(0.075)
	[0.300]	[0.036]	[0.162]	[0.028]
Test $NL = L$, p-value	0.00	0.36	0.28	0.72
Panel D. Subtreatments by Types of Letters				
$[NL]$ No Letter - Any Ad	0.217***	0.484***	0.115**	0.212**
	(0.059)	(0.172)	(0.051)	(0.099)
	[0.000]	[0.002]	[0.018]	[0.008]
$[PL]$ Partial Knowledge Letter - Any Ad	0.007	0.100	0.030	0.038
	(0.057)	(0.163)	(0.048)	(0.064)
	[0.894]	[0.528]	[0.546]	[0.582]
$[FL]$ Full Knowledge Letter - Any Ad	0.090	0.530**	0.092*	0.307***
	(0.059)	(0.220)	(0.052)	(0.118)
	[0.140]	[0.004]	[0.078]	[0.002]
Test $NL = PL$, p-value	0.00	0.04	0.12	0.08
Test $NL = FL$, p-value	0.05	0.85	0.70	0.50
Test $PL = FL$, p-value	0.20	0.07	0.27	0.02
Control Mean	0.29	0.55	0.16	0.20
Sample Size	677	677	677	677

Notes: This table reports the same estimates as in Table 2 without any control variables except for strata fixed effects. The outcome in column (1) is an indicator for whether any report was issued to the MOE from each municipality. In column (2) it is the number of such reports. In columns (3)-(4) the same definitions are used on the subset of reports of a high quality (see Section 2 for a discussion about how quality of reports is assessed by the MOE). All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value $<$ 0.01, ** p-value $<$ 0.05, * p-value $<$ 0.1.

Table A15: Robustness: Impacts on the Media-Based Irregularity Measures - No Controls

	(1) Media Irregularities (=1)	(2) Number of Media Irregularities
Panel A. Pooled Treatment		
[T] Any treatment	-0.075**	-0.094**
	(0.033)	(0.043)
	[0.002]	[0.014]
Panel B. Subtreatments by Types of Ads		
[IA] Information Ad	-0.061	-0.089*
	(0.041)	(0.050)
	[0.160]	[0.098]
[CA] Call-to-Action Ad	-0.048	-0.057
	(0.042)	(0.057)
	[0.268]	[0.324]
[$I + CA$] Info + Call-to-Action Ad	-0.117***	-0.136***
	(0.033)	(0.044)
	[0.000]	[0.006]
Test $IA = CA$, p-value	0.74	0.54
Test $IA = I + CA$, p-value	0.08	0.22
Test $CA = I + CA$, p-value	0.04	0.10
Panel C. Subtreatments by Letter - No Letter		
[NL] No Letter - Any Ad	-0.059	-0.072
	(0.040)	(0.054)
	[0.140]	[0.204]
[L] Letter - Any Ad	-0.084**	-0.106**
	(0.033)	(0.043)
	[0.004]	[0.006]
Test $NL = L$, p-value	0.42	0.41
Panel D. Subtreatments by Types of Letters		
[NL] No Letter - Any Ad	-0.059	-0.072
	(0.040)	(0.054)
	[0.150]	[0.212]
[PL] Partial Knowledge Letter - Any Ad	-0.070*	-0.102**
	(0.040)	(0.046)
	[0.074]	[0.050]
[FL] Full Knowledge Letter - Any Ad	-0.098***	-0.109**
	(0.035)	(0.050)
	[0.018]	[0.028]
Test $NL = PL$, p-value	0.77	0.50
Test $NL = FL$, p-value	0.24	0.43
Test $PL = FL$, p-value	0.43	0.87
Control Mean	0.14	0.17
Sample Size	677	677

Notes: This table reports the same estimates as in Table 2 without any control variables except for strata fixed effects. The outcome in column (1) is an indicator for whether any irregularity was reported in the media in a particular municipality. In column (2) it is the number of different irregularities. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value $<$ 0.05, * p-value < 0.1.

Table A16: Robustness: Impacts on Deviations from Benford's Second Digit Law - No Controls

	(1) Index of all Forensic Test Stats (z-score)	(2) Any P-value From Forensic Tests < 0.1 (=1)	(3) Any P-value From Forensic Tests < 0.05 (=1)
Panel A. Pooled Treatment			
$[T]$ Any treatment	-0.158*	-0.105**	-0.119**
	(0.091)	(0.045)	(0.047)
	[0.120]	[0.028]	[0.018]
Panel B. Subtreatments by Types of Ads			
$[IA]$ Information Ad	-0.229**	-0.139**	-0.140**
	(0.117)	(0.059)	(0.060)
	[0.048]	[0.018]	[0.014]
$[CA]$ Call-to-Action Ad	-0.083	-0.046	-0.063
	(0.117)	(0.059)	(0.061)
	[0.482]	[0.448]	[0.288]
$[I+CA]$ Info + Call-to-Action Ad	-0.160	-0.129**	-0.152**
	(0.112)	(0.059)	(0.060)
	[0.150]	[0.018]	[0.020]
Test $IA = CA$, p-value	0.24	0.16	0.24
Test $IA = I + CA$, p-value	0.57	0.88	0.86
Test $CA = I + CA$, p-value	0.52	0.21	0.18
Panel C. Subtreatments by Letter - No Letter			
$[NL]$ No Letter - Any Ad	-0.122	-0.068	-0.099*
	(0.115)	(0.058)	(0.060)
	[0.310]	[0.230]	[0.114]
$[L]$ Letter - Any Ad	-0.176*	-0.124**	-0.129**
	(0.098)	(0.049)	(0.051)
	[0.090]	[0.012]	[0.008]
Test $NL = L$, p-value	0.61	0.32	0.60
Panel D. Subtreatments by Types of Letters			
$[NL]$ No Letter - Any Ad	-0.122	-0.068	-0.099*
	(0.115)	(0.058)	(0.060)
	[0.278]	[0.228]	[0.118]
$[PL]$ Partial Knowledge Letter - Any Ad	-0.241**	-0.145**	-0.143**
	(0.108)	(0.060)	(0.061)
	[0.030]	[0.018]	[0.010]
$[FL]$ Full Knowledge Letter - Any Ad	-0.112	-0.104*	-0.115*
	(0.122)	(0.059)	(0.061)
	[0.326]	[0.074]	[0.040]
Test $NL = PL$, p-value	0.30	0.25	0.51
Test $NL = FL$, p-value	0.93	0.58	0.81
Test $PL = FL$, p-value	0.29	0.54	0.68
Control Mean	0.00	0.60	0.52
Sample Size	677	677	677

Notes: This table reports the same estimates as in Table 2 without any control variables except for strata fixed effects. The outcome in column (1) is the index of the χ^2, Kolmogorov-Smirnov and Kuiper test statistics testing for Benford's 2nd digit law, described in Section 4.2.2 . In columns (2) and (3) it is an indicator that takes the value of one if the p-value of any of these tests leads to rejection of the null hypothesis with less than a 10% or 5% significance level. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A17: Robustness: Impacts on Vote Share of Candidates Likely to Engage in Irregularities - No Controls

	(1)	(2)	(3)	(4)
		Vote Share (%)		
Interaction term Z:	Candidate will engage in irregularities (fraction of respondents)	Demeaned Candidate will engage in irregularities (fraction of respondents)	Above Average Candidate will engage in irregularities (=1)	Past Malfeasance (=1)
Panel A. Pooled Treatment				
$[T \times Z]$ Any treatment $\times Z$	-3.852	-6.728***	-3.942***	-7.417*
	(2.562)	(2.601)	(1.468)	(4.296)
	[0.078]	[0.012]	[0.002]	[0.088]
	{0.138}	{0.006}	{0.007}	
Panel B. Subtreatments by Types of Ads				
$[IA \times Z]$ Information Ad $\times Z$	-4.049	-8.493***	-3.845**	-13.503***
	(3.033)	(3.078)	(1.846)	(3.640)
	[0.186]	[0.016]	[0.042]	[0.006]
	{0.196}	{0.005}	{0.033}	
$[CA \times Z]$ Call-to-Action Ad $\times Z$	-2.572	-2.448	-2.548	-0.020
	(3.572)	(3.660)	(2.002)	(8.601)
	[0.518]	[0.572]	[0.208]	[1.000]
	{0.474}	{0.518}	{0.205}	
$[I + CA \times Z]$ Info + Call-to-Action Ad $\times Z$	-4.931	-8.799**	-5.507***	-3.625
	(3.435)	(3.659)	(1.864)	(5.811)
	[0.146]	[0.032]	[0.004]	[0.614]
	{0.169}	{0.018}	{0.003}	
Test $IA \times Z = CA \times Z$, p-value	0.68	0.12	0.55	0.13
Test $IA \times Z = I + CA \times Z$, p-value	0.80	0.94	0.41	0.09
Test $CA \times Z = I + CA \times Z$, p-value	0.56	0.14	0.17	0.71
Panel C. Subtreatments by Letter - No Letter				
$[NL \times Z]$ No Letter - Any Ad $\times Z$	-5.538*	-7.011**	-3.505*	-3.593
	(3.074)	(3.093)	(1.900)	(5.129)
	[0.082]	[0.030]	[0.066]	[0.542]
	{0.073}	{0.019}	{0.067}	
$[L \times Z]$ Letter - Any Ad $\times Z$	-2.833	-6.565**	-4.205***	-10.433*
	(2.847)	(2.992)	(1.599)	(5.433)
	[0.308]	[0.016]	[0.014]	[0.080]
	{0.341}	{0.031}	{0.008}	
Test $NL \times Z = L \times Z$, p-value	0.36	0.89	0.70	0.29
Panel D. Subtreatments by Types of Letters				
$[NL \times Z]$ No Letter - Any Ad $\times Z$	-5.543*	-7.011**	-3.505*	-3.506
	(3.074)	(3.094)	(1.901)	(5.146)
	[0.072]	[0.044]	[0.054]	[0.564]
	{0.075}	{0.020}	{0.068}	
$[FL \times Z]$ Full Knowledge Letter - Any Ad $\times Z$	-3.310	-7.267**	-4.474**	-4.734
	(3.658)	(3.691)	(2.006)	(9.494)
	[0.336]	[0.060]	[0.018]	[0.510]
	{0.389}	{0.045}	{0.028}	
$[PL \times Z]$ Partial Knowledge Letter - Any Ad $\times Z$	-2.236	-5.760	-3.946**	-14.983***
	(3.211)	(3.776)	(1.834)	(5.472)
	[0.504]	[0.114]	[0.028]	[0.036]
	{0.493}	{0.138}	{0.032}	
Test $NL \times Z = PL \times Z$, p-value	0.32	0.75	0.83	0.08
Test $NL \times Z = FL \times Z$, p-value	0.55	0.95	0.66	0.90
Test $PL \times Z = FL \times Z$, p-value	0.78	0.73	0.80	0.32
Control Mean	18.91	18.91	18.91	18.91
Sample Size	2986	2986	2986	263
N. Municipalities	628	628	628	48

Notes: This table reports the same estimates as in Table 5 without any control variables except for strata fixed effects. The outcome in all columns is the vote share for each candidate, expressed as a percentage of total valid votes. In each of these columns, a different measure of the likelihood that a candidate commits irregularities is used to compute the candidate-level heterogenous effects. In column (1) it is the proportion of respondents from the pre-treatment survey that say that the candidate might commit at least one type of electoral irregularity. In column (2) the outcome is this same variable, demeaned using the municipality-level mean. In column (3) it is an indicator that takes the value of one if this variable is above the municipal-level mean. Finally, in column (4) it is an indicator for whether a candidate was found to be involved in malfeasance in the past according to the investigation by the NGO PARES. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Clustered standard errors at the municipality level are shown in parentheses, random inference p-values are shown in square brackets, and clustered wild-bootstrap p-values correcting for the variance in estimating Z are shown in curly brackets; *** p-value $<$ 0.01, ** p-value $<$ 0.05, * p-value $<$ 0.1.

Table A18: Robustness: Impacts on Additional Electoral Outcomes - No Controls

	(1) Turnout (%)	(2) Margin of Victory (%)
Panel A. Pooled Treatment		
[T] Any treatment	-0.507	1.002
	(0.733)	(0.974)
	[0.494]	[0.376]
Panel B. Subtreatments by Types of Ads		
[IA] Information Ad	0.123	1.211
	(0.959)	(1.448)
	[0.900]	[0.394]
[CA] Call-to-Action Ad	-0.374	1.384
	(0.837)	(1.216)
	[0.690]	[0.264]
[$I + CA$] Info + Call-to-Action Ad	-1.269	0.417
	(1.065)	(1.251)
	[0.210]	[0.726]
Test $IA = CA$, p-value	0.61	0.91
Test $IA = I + CA$, p-value	0.23	0.62
Test $CA = I + CA$, p-value	0.40	0.48
Panel C. Subtreatments by Letter - No Letter		
[NL] No Letter - Any Ad	0.153	0.686
	(0.852)	(1.260)
	[0.848]	[0.618]
[L] Letter - Any Ad	-0.851	1.167
	(0.830)	(1.082)
	[0.352]	[0.364]
Test $NL = L$, p-value	0.24	0.70
Panel D. Subtreatments by Types of Letters		
[NL] No Letter - Any Ad	0.153	0.686
	(0.852)	(1.260)
	[0.856]	[0.560]
[PL] Partial Knowledge Letter - Any Ad	-0.706	0.103
	(1.066)	(1.222)
	[0.512]	[0.926]
[FL] Full Knowledge Letter - Any Ad	-0.995	2.223
	(0.936)	(1.427)
	[0.266]	[0.082]
Test $NL = PL$, p-value	0.43	0.67
Test $NL = FL$, p-value	0.23	0.33
Test $PL = FL$, p-value	0.80	0.17
Control Mean	67.10	12.03
Sample Size	677	677

Notes: This table reports the same estimates as in Table 6 without any control variables except for strata fixed effects. The outcome in column (1) is turnout, as a percentage of the people registered to vote. In column (2) it is the margin of the winning candidate over the runner-up , expressed as a percentage of total votes. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A19: Robustness: Impacts on Reports - No Weights

	(1) Reports (=1)	(2) N. Reports	(3) High Quality Reports (=1)	(4) High Quality N. Reports
Panel A. Pooled Treatment				
[T] Any treatment	0.106***	0.366***	0.083***	0.173***
	(0.035)	(0.101)	(0.030)	(0.054)
	[0.004]	[0.002]	[0.012]	[0.000]
Panel B. Subtreatments by Types of Ads				
[IA] Information Ad	0.145***	0.452***	0.069*	0.114*
	(0.048)	(0.147)	(0.040)	(0.064)
	[0.000]	[0.000]	[0.056]	[0.062]
[CA] Call-to-Action Ad	0.016	0.180	0.069*	0.144*
	(0.046)	(0.137)	(0.041)	(0.083)
	[0.750]	[0.170]	[0.070]	[0.058]
[$I+CA$] Info + Call-to-Action Ad	0.155***	0.462***	0.109***	0.259***
	(0.046)	(0.148)	(0.041)	(0.086)
	[0.000]	[0.002]	[0.016]	[0.002]
Test $IA = CA$, p-value	0.02	0.12	0.99	0.74
Test $IA = I + CA$, p-value	0.84	0.96	0.39	0.13
Test $CA = I + CA$, p-value	0.01	0.11	0.39	0.29
Panel C. Subtreatments by Letter - No Letter				
[NL] No Letter - Any Ad	0.216***	0.505***	0.138***	0.246***
	(0.048)	(0.135)	(0.043)	(0.084)
	[0.000]	[0.000]	[0.002]	[0.000]
[L] Letter - Any Ad	0.049	0.296**	0.054*	0.136**
	(0.038)	(0.115)	(0.032)	(0.058)
	[0.198]	[0.020]	[0.098]	[0.024]
Test $NL = L$, p-value	0.00	0.16	0.04	0.21
Panel D. Subtreatments by Types of Letters				
[NL] No Letter - Any Ad	0.216***	0.506***	0.138***	0.247***
	(0.048)	(0.135)	(0.043)	(0.084)
	[0.000]	[0.000]	[0.002]	[0.000]
[PL] Partial Knowledge Letter - Any Ad	0.035	0.167	0.032	0.047
	(0.045)	(0.128)	(0.038)	(0.058)
	[0.450]	[0.176]	[0.392]	[0.414]
[FL] Full Knowledge Letter - Any Ad	0.063	0.423**	0.076*	0.225**
	(0.046)	(0.165)	(0.040)	(0.087)
	[0.178]	[0.004]	[0.058]	[0.002]
Test $NL = PL$, p-value	0.00	0.03	0.02	0.02
Test $NL = FL$, p-value	0.00	0.67	0.19	0.84
Test $PL = FL$, p-value	0.59	0.17	0.32	0.05
Control Mean	0.29	0.55	0.16	0.20
Sample Size	698	698	698	698

Notes: This table reports the same estimates as in Table 2 without using the weights described in Section 3.6. The outcome in column (1) is an indicator for whether any report was issued to the MOE from each municipality. In column (2) it is the number of such reports. In columns (3)-(4) the same definitions are used on the subset of reports of a high quality (see Section 2 for a discussion about how quality of reports is assessed by the MOE). All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A20: Robustness: Impacts on the Media-Based Irregularity Measures - No Weights

	(1) Media Irregularities (=1)	(2) Number of Media Irregularities
Panel A. Pooled Treatment		
[T] Any treatment	-0.052**	-0.066*
	(0.026)	(0.035)
	[0.040]	[0.046]
Panel B. Subtreatments by Types of Ads		
[IA] Information Ad	-0.050	-0.065
	(0.033)	(0.041)
	[0.122]	[0.166]
[CA] Call-to-Action Ad	-0.043	-0.062
	(0.033)	(0.042)
	[0.184]	[0.118]
[$I+CA$] Info + Call-to-Action Ad	-0.062**	-0.072*
	(0.031)	(0.042)
	[0.048]	[0.114]
Test $IA = CA$, p-value	0.84	0.95
Test $IA = I + CA$, p-value	0.71	0.84
Test $CA = I + CA$, p-value	0.56	0.80
Panel C. Subtreatments by Letter - No Letter		
[NL] No Letter - Any Ad	-0.030	-0.037
	(0.033)	(0.044)
	[0.350]	[0.426]
[L] Letter - Any Ad	-0.063**	-0.082**
	(0.027)	(0.036)
	[0.014]	[0.022]
Test $NL = L$, p-value	0.26	0.22
Panel D. Subtreatments by Types of Letters		
[NL] No Letter - Any Ad	-0.030	-0.037
	(0.033)	(0.044)
	[0.362]	[0.424]
[PL] Partial Knowledge Letter - Any Ad	-0.070**	-0.100***
	(0.031)	(0.038)
	[0.028]	[0.016]
[FL] Full Knowledge Letter - Any Ad	-0.057*	-0.063
	(0.032)	(0.043)
	[0.086]	[0.148]
Test $NL = PL$, p-value	0.23	0.09
Test $NL = FL$, p-value	0.44	0.54
Test $PL = FL$, p-value	0.67	0.30
Control Mean	0.14	0.17
Sample Size	698	698

Notes: This table reports the same estimates as in Table 3 without using the weights described in Section 3.6. The outcome in column (1) is an indicator for whether any irregularity was reported in the media in a particular municipality. In column (2) it is the number of different irregularities. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A21: Robustness: Impacts on Deviations from Benford's Second Digit Law - No Weights

	(1) Index of all Forensic Test Stats (z-score)	(2) Any P-value From Forensic Tests < 0.1 (=1)	(3) Any P-value From Forensic Tests < 0.05 (=1)
Panel A. Pooled Treatment			
$[T]$ Any treatment	-0.137**	-0.059	-0.086**
	(0.069)	(0.038)	(0.038)
	[0.054]	[0.134]	[0.036]
Panel B. Subtreatments by Types of Ads			
$[IA]$ Information Ad	-0.258***	-0.098**	-0.124**
	(0.085)	(0.049)	(0.049)
	[0.002]	[0.046]	[0.006]
$[CA]$ Call-to-Action Ad	-0.104	-0.053	-0.078
	(0.087)	(0.051)	(0.051)
	[0.236]	[0.284]	[0.118]
$[I+CA]$ Info + Call-to-Action Ad	-0.048	-0.026	-0.057
	(0.091)	(0.049)	(0.048)
	[0.612]	[0.628]	[0.208]
Test $IA = CA$, p-value	0.09	0.41	0.39
Test $IA = I + CA$, p-value	0.02	0.17	0.19
Test $CA = I + CA$, p-value	0.55	0.63	0.70
Panel C. Subtreatments by Letter - No Letter			
$[NL]$ No Letter - Any Ad	-0.059	-0.044	-0.076
	(0.088)	(0.049)	(0.050)
	[0.510]	[0.390]	[0.122]
$[L]$ Letter - Any Ad	-0.177**	-0.066	-0.091**
	(0.074)	(0.041)	(0.041)
	[0.024]	[0.106]	[0.028]
Test $NL = L$, p-value	0.14	0.62	0.75
Panel D. Subtreatments by Types of Letters			
$[NL]$ No Letter - Any Ad	-0.059	-0.044	-0.076
	(0.088)	(0.049)	(0.050)
	[0.490]	[0.386]	[0.144]
$[PL]$ Partial Knowledge Letter - Any Ad	-0.197**	-0.081	-0.100**
	(0.081)	(0.050)	(0.050)
	[0.018]	[0.108]	[0.044]
$[FL]$ Full Knowledge Letter - Any Ad	-0.157*	-0.053	-0.082*
	(0.092)	(0.049)	(0.049)
	[0.070]	[0.290]	[0.078]
Test $NL = PL$, p-value	0.12	0.50	0.65
Test $NL = FL$, p-value	0.32	0.87	0.91
Test $PL = FL$, p-value	0.66	0.61	0.73
Control Mean	0.00	0.60	0.52
Sample Size	698	698	698

Notes: This table reports the same estimates as in Table 4 without using the weights described in Section 3.6. The outcome in column (1) is the index of the χ^2, Kolmogorov-Smirnov and Kuiper test statistics testing for Benford's 2nd digit law, described in Section 4.2.2 . In columns (2) and (3) it is an indicator that takes the value of one if the p-value of any of these tests leads to rejection of the null hypothesis with less than a 10% or 5% significance level. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A22: Robustness: Impacts on Vote Share of Candidates Likely to Engage in Irregularities - No Weights

	(1)	(2)	(3)	(4)
	Vote Share (%)			
Interaction term Z:	Candidate will engage in irregularities (fraction of respondents)	Demeaned Candidate will engage in irregularities (fraction of respondents)	Above Average Candidate will engage in irregularities (=1)	Past Malfeasance (=1)
Panel A. Pooled Treatment				
$[T \times Z]$ Any treatment $\times Z$	-3.444*	-3.656	-2.635**	-4.237
	(1.823)	(2.353)	(1.217)	(4.014)
	[0.044]	[0.104]	[0.012]	[0.216]
	{0.062}	{0.122}	{0.030}	
Panel B. Subtreatments by Types of Ads				
$[IA \times Z]$ Information Ad $\times Z$	-3.724	-4.065	-1.895	-10.801***
	(2.394)	(3.146)	(1.586)	(3.415)
	[0.136]	[0.208]	[0.246]	[0.024]
	{0.133}	{0.215}	{0.235}	
$[CA \times Z]$ Call-to-Action Ad $\times Z$	-1.188	-1.273	-1.826	1.203
	(2.403)	(3.062)	(1.601)	(6.735)
	[0.620]	[0.728]	[0.270]	[0.858]
	{0.615}	{0.675}	{0.246}	
$[I + CA \times Z]$ Info + Call-to-Action Ad $\times Z$	-5.147**	-5.526*	-4.231***	-0.301
	(2.290)	(3.040)	(1.551)	(4.736)
	[0.024]	[0.086]	[0.008]	[0.954]
	{0.025}	{0.071}	{0.006}	
Test $IA \times Z = CA \times Z$, p-value	0.34	0.43	0.97	0.07
Test $IA \times Z = I + CA \times Z$, p-value	0.58	0.68	0.18	0.01
Test $CA \times Z = I + CA \times Z$, p-value	0.12	0.21	0.17	0.83
Panel C. Subtreatments by Letter - No Letter				
$[NL \times Z]$ No Letter - Any Ad $\times Z$	-3.139	-2.766	-1.785	-4.779
	(2.261)	(2.920)	(1.586)	(4.498)
	[0.172]	[0.356]	[0.264]	[0.348]
	{0.174}	{0.344}	{0.261}	
$[L \times Z]$ Letter - Any Ad $\times Z$	-3.607*	-4.162	-3.087**	-3.933
	(2.001)	(2.628)	(1.315)	(4.990)
	[0.072]	[0.112]	[0.016]	[0.480]
	{0.078}	{0.120}	{0.022}	
Test $NL \times Z = L \times Z$, p-value	0.83	0.63	0.39	0.87
Panel D. Subtreatments by Types of Letters				
$[NL \times Z]$ No Letter - Any Ad $\times Z$	-3.139	-2.766	-1.785	-4.528
	(2.261)	(2.921)	(1.587)	(4.478)
	[0.172]	[0.396]	[0.254]	[0.364]
	{0.173}	{0.346}	{0.264}	
$[FL \times Z]$ Full Knowledge Letter - Any Ad $\times Z$	-2.449	-3.092	-2.181	3.139
	(2.467)	(3.132)	(1.594)	(6.718)
	[0.300]	[0.360]	[0.178]	[0.596]
	{0.333}	{0.323}	{0.168}	
$[PL \times Z]$ Partial Knowledge Letter - Any Ad $\times Z$	-4.798**	-5.431*	-3.986**	-11.353**
	(2.350)	(3.283)	(1.560)	(4.807)
	[0.034]	[0.080]	[0.006]	[0.070]
	{0.047}	{0.108}	{0.012}	
Test $NL \times Z = PL \times Z$, p-value	0.51	0.45	0.20	0.18
Test $NL \times Z = FL \times Z$, p-value	0.79	0.92	0.82	0.27
Test $PL \times Z = FL \times Z$, p-value	0.38	0.53	0.30	0.05
Control Mean	18.91	18.91	18.91	18.91
Sample Size	2989	2989	2989	263
N. Municipalities	630	630	630	48

Notes: This table reports the same estimates as in Table 5 without using the weights described in Section 3.6. The outcome in all columns is the vote share for each candidate, expressed as a percentage of total valid votes. In each of these columns, a different measure of the likelihood that a candidate commits irregularities is used to compute the candidate-level heterogenous effects. In column (1) it is the proportion of respondents from the pre-treatment survey that say that the candidate might commit at least one type of electoral irregularity. In column (2) the outcome is this same variable, demeaned using the municipality-level mean. In column (3) it is an indicator that takes the value of one if this variable is above the municipal-level mean. Finally, in column (4) it is an indicator for whether a candidate was found to be involved in malfeasance in the past according to the investigation by the NGO PARES. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Clustered standard errors at the municipality level are shown in parentheses, random inference p-values are shown in square brackets, and clustered wild-bootstrap p-values correcting for the variance in estimating Z are shown in curly brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A23: Robustness: Impacts on Additional Electoral Outcomes - No Weights

	(1) Turnout (%)	(2) Margin of Victory (%)
Panel A. Pooled Treatment		
[T] Any treatment	0.165	0.817
	(0.464)	(0.829)
	[0.696]	[0.354]
Panel B. Subtreatments by Types of Ads		
[IA] Information Ad	-0.122	1.455
	(0.605)	(1.194)
	[0.838]	[0.180]
[CA] Call-to-Action Ad	0.461	1.103
	(0.523)	(1.040)
	[0.466]	[0.296]
[$I + CA$] Info + Call-to-Action Ad	0.160	-0.108
	(0.642)	(1.070)
	[0.820]	[0.920]
Test $IA = CA$, p-value	0.32	0.78
Test $IA = I + CA$, p-value	0.68	0.23
Test $CA = I + CA$, p-value	0.63	0.30
Panel C. Subtreatments by Letter - No Letter		
[NL] No Letter - Any Ad	-0.206	0.568
	(0.629)	(1.096)
	[0.742]	[0.614]
[L] Letter - Any Ad	0.357	0.946
	(0.485)	(0.908)
	[0.440]	[0.332]
Test $NL = L$, p-value	0.33	0.73
Panel D. Subtreatments by Types of Letters		
[NL] No Letter - Any Ad	-0.206	0.568
	(0.630)	(1.097)
	[0.756]	[0.566]
[PL] Partial Knowledge Letter - Any Ad	0.325	0.449
	(0.553)	(1.103)
	[0.598]	[0.688]
[FL] Full Knowledge Letter - Any Ad	0.389	1.437
	(0.584)	(1.111)
	[0.538]	[0.164]
Test $NL = PL$, p-value	0.40	0.92
Test $NL = FL$, p-value	0.37	0.49
Test $PL = FL$, p-value	0.91	0.43
Control Mean	67.10	12.03
Sample Size	698	697

Notes: This table reports the same estimates as in Table 6 without using the weights described in Section 3.6. The outcome in column (1) is turnout, as a percentage of the people registered to vote. In column (2) it is the margin of the winning candidate over the runner-up , expressed as a percentage of total votes. All specifications include the covariates selected using the method method described in Chernozhukov et al., 2015 and Belloni et al., 2014. Robust standard errors are shown in parentheses and random inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

A Dataset of News about Electoral Irregularities

We constructed our dataset of the electoral irregularities covered in news from three different and complementary sources. We begin by discussing each of these sources and then we describe further details about the coding of news.

1.Private News Monitoring Company: We hired *Siglo Data*, a data analysis firm based in Colombia, which is specialized in monitoring and classifying news in mass media and social networks. They actively monitor and classify news coming from TV channels, radio stations, written press, both offline and online. More concretely, they claim to cover news from over 100 newspapers and 60 magazines, 10 national and 60 radiostations, 15 national and 10 regional TV channels and 370 news websites. They provided the data for the universe of news that had mentions about electoral irregularities related to the 2019 local elections, and that appeared in either written press, the radio or newspapers on the internet up to November 10, two weeks after the date of elections.

2. Press releases from Colombia's Electoral Court: We additionally relied on the official press releases produced by the *Consejo Nacional Electoral* (CNE), the National Electoral Court of Colombia. As part of its regular activities, this entity is in charge of monitoring and controlling all electoral activity of political groups and their candidates, as well as overseeing the electoral organization and ensuring the proper development of electoral processes. In this vein, the CNE published on its website a list of news covering the 2019 local elections, which we included in our analysis.[42]

3. Electoral irregularity monitoring from an NGO: Finally, we use information from the *Fundación Paz y Reconciliación* (PARES), a well-known NGO in Colombia focused on producing independent research on conflict, security, governance, democracy, among others. To this end, they monitor elections, and in 2019 they produced a report of instances of electoral irregularities they gathered from media, citizen reports and their own monitoring activity in the field.[43]

Using the exhaustive list of news coming from these three sources, we hired a research assistant to go through each of these and classify it in terms of three main variables: (i) whether the news reported an electoral irregularity or not, (ii) whether the news came from information coming from citizen reports submitted to the MOE, (iii) the types of electoral irregularities reported in the news. In our final data set, we exclude news that that are not

[42]The website they use to publicize the covarage of elections can be accessed through the following link:https://www.cne.gov.co/prensa/cne-en-medios?start=78.

[43]The full report is publicly available online at this link: https://pares.com.co/2019/10/29/un-balance-nacional-de-estas-elecciones-locales-2019/.

about electoral irregularities according to (i), and those in (ii) so that we do not confound the effect of the campaign on reporting from the effect on actual irregularities.

Using the cleaned data set, which contains over 160 news satisfying our criteria, we create an indicator of whether any electoral irregularity was reported by the media in each of one the municipalities in our study group, as well as a variable of the the number of distinct irregularities reported per municipality.

B Covariates Included in the Analysis

The covariates included in the analysis can be broadly categorized in three groups:

Past reports: We include the number of reports made to the MOE in the 2015 local elections and the 2018 congressional elections, as a way to control both for previous experience with reporting channels and the prevalence of electoral irregularities.

Socioeconomic characteristics: As geographical and demographic variables we use the municipal population in 2018, the proportion of rural population and dummies for the six main regions in the country. As measures of economic activity and development we use GDP per capita and the % of poor population. All of these variables were taken from the the National Department of Statistics (DANE) except for the last one which was taken from the "Muncipal Characteristics" database created and updated by the Centro de Estudios sobre Desarrollo Económico (CEDE), at the Universidad de los Andes in Bogotá, Colombia. From this database we also used the municipal homicide rate as a proxy for local violence. Finally, we also included Facebook's penetration rate (defined as the number of active Facebook users divided by total population) which we construct using user data from Facebook's Marketing API.

Political preferences: In order to get a rich set of political characteristics for each municipality we used the turnout, the margin of victory and the share of blank votes in the 2015 mayoral elections, as well as the vote share for each major party in the 2018 congressional elections and the vote share for each candidate in the second round of the 2014 presidential elections. Lastly, we include the number of candidates running in the 2019 mayoral elections in the analysis. All of these variables were constructed from the official voting records held by the Registraduría Nacional.

When running regressions at the candidate-level, we also included the following covariates for each of the mayoral candidates running in the municipalities in our sample:

Political platform: We included information about the political platform used by the candidate to register her candidacy. Specifically, whether she is running with a coalition of parties or as an independent. We also created a variable of whether the candidate is running as an incumbent or not. In Colombia, there is not immediate reelection for candidates. Thus, we computed being an incumbent as either running with exactly the same party or coalition of parties as the incumbent mayor (strict measure) or with at least one party shared by the party or coalition of parties as the incumbent mayor (lax measure).

Sociodemographic characteristics: We used data on the candidates' gender and age, which was provided by the Registraduría Nacional.

C Measures of Deviation from Benford's 2nd Digit Law

We use the following tests to determine whether the observed distribution of second digits in the voting booth counts differs from Benford's distribution in each municipality in our sample:

1. Pearson X^2:

$$X^2 = n \times \sum_{i=0}^{9} \frac{(o_i - e_i)^2}{e_i}$$

   where o_i is the observed proportion of digit i and e_i is the expected proportion according to Benford's distribution. This statistic is distributed χ^2 with 9 degrees of freedom.

2. Kolmogorov-Smirnov D:

$$D = \sup_{x \in \{0,1,2,\ldots,9\}} |F_n(x) - F(x)|$$

   where F_n is the empirical distribution of digits and $F(x)$ is the target distribution.

3. Kuiper V:

$$V = \max_{x \in \{0,1,2,\ldots,9\}} [F_n(x) - F(x)] + \max_{x \in \{0,1,2,\ldots,9\}} [F(x) - F_n(x)]$$

D Bootstrap Procedure to Account for Variance in Estimating Candidate-Level Variables

Denote the data by the triplet (y, X, Z), where y and X are at the candidate-municipality level, while Z corresponds to data from our pre-treatment survey, and thus is at the at the respondent-candidate-municipality level. The candidate level variables $\hat{z}_{ic}$ are estimated from Z but are included as regressors in the estimation so $\hat{z}_{ic} \subset X$.

Given this notation, we we build on the wild bootstrap procedure proposed by Cameron et al. (2008), and incorporate an extra resampling stage of Z, in order to incorporate the variance in estimating the candidate level variables $\hat{z}_{ic}$. The procedure can be summarized as follows:

1. From the original sample, estimate $t = \frac{\hat{\gamma}-\gamma_0}{s_{\hat{\gamma}}}$, where $s_{\hat{\gamma}}$ is the standard error of $\hat{\gamma}$ clustered at the municipality level.

2. Estimate the restricted model which imposes the null hypothesis (i.e. $\gamma = 0$). Call the restricted estimates $\hat{\beta}^R$ and the corresponding residuals $\left\{(\hat{u}_1^R, ..., \hat{u}_M^R)\right\}$.

3. Do B iterations of this step. On the $b - th$ iteration:

   (a) For each candidate-municipality combination, create a sample of respondents Z^*_{ci} by resampling with replacement N_{ic} times from the original sample of respondents—where N_{ic} is the original number of respondents for candidate c in municipality i .

   (b) Compute the $\hat{z}^*_{ic}$, the measure of how likely each candidate is to engage in irregularities from Z^*.

   (c) Create a pseudo-sample (y^*, X^*) using the following method. For each cluster $i = 1, ..., M$, generate $\hat{u}_i^{R*} = a_i \hat{u}_i^R$, where a_i is a random variable that takes the value $\frac{1-\sqrt{5}}{2}$ with probability $\frac{1+\sqrt{5}}{2\sqrt{5}}$, or $1 - \frac{1-\sqrt{5}}{2}$ with probability $1 - \frac{1+\sqrt{5}}{2\sqrt{5}}$.[44] Then define X^* as X but using the sampled $\hat{z}^*_{ic}$ instead of the original $\hat{z}_{ic}$. Finally, define $y_i^* = X_i^* \hat{\beta}^R + \hat{u}_i^{R*}$

   (d) Compute $t_b = \frac{\hat{\gamma}_b^* - \gamma_0}{s_{\hat{\gamma}_b^*}}$, where $\hat{\gamma}_b^*$ and $s_{\hat{\gamma}_b^*}$ are estimated as in step (1) but using the $b - th$ pseudo-sample.

4. Compute the bootstrapped p-value as $p = \frac{\sum_{b=1}^{B} \mathbb{1}\{|t_b| > |t|\}}{B}$

[44] As explained by Cameron et al. (2008) these weights are preferred when the distribution of the estimates is potentially asymmetric. We use this alternative since it produces the most conservative p-values (i.e. the ones most likely not to reject the null) when applied to our setting.