Electoral Integrity and Local Development: Experimental Evidence From Colombia

Mateo Montenegro and Natalia Garbiras-Díaz*

July 22, 2026

Abstract

Today most countries hold elections to choose their leaders, yet only about half of these elections are free and fair. This paper studies whether enhancing electoral integrity improves the provision of basic public services, a central dimension of local development. We leverage experimental variation from a randomized intervention during Colombia's 2019 mayoral elections, which successfully reduced irregularities such as fraud and vote buying across nearly 700 municipalities. We document four sets of results. First, municipalities with cleaner elections experienced increased access to local public services over the following four years. Selection and accountability channels underpinning this result, and develop a framework to quantify the contribution of each channel. We find that selection accounts for 22–30 percent of the effect, while accountability explains the remainder. Third, we identify specific mayoral actions behind these gains: local governments in treated municipalities adopted more competitive procurement practices and faced fewer corruption-related judicial investigations. Finally, we rule out alternative mechanisms, including increased fiscal capacity and support from higher levels of government. Taken together, the results show that safeguarding democratic processes yields tangible welfare gains for citizens.

JEL Codes: C93, D02, D72, D73, H41, H70, O12, O43.

*Montenegro: Toulouse School of Economics (email: mateo.montenegro@tse-fr.eu); Garbiras-Díaz: Harvard Business School (email: ngarbirasdiaz@hbs.edu). We thank Hernán Carvajal, Alejandro Castillo, Juan Gómez, Felipe Junco, Konstantinos Mira and Paula Remicio for their outstanding research assistance. We are grateful for the valuable comments and suggestions received from participants at the BREAD, CEPR Development, EPSA, EUDN and GATE conferences, as well as from seminar audiences at Brown University, CEMFI, CIPR, Columbia University, CUNEF, FMI at Texas Tech, NYU, PSE, TSE, and the World Bank. We are especially grateful to Horacio Larreguy and Vincent Pons for their helpful comments on this manuscript. The original experiment was approved by MIT's IRB (the *Committee on the Use of Humans as Experimental Subjects*) with reference #1904805455 and it was registered in the AEA RCT Registry with unique identifying number: https://doi.org/10.1257/rct.4678-1.0. Mateo Montenegro acknowledges funding from ANR under grant ANR-17-EURE-0010 for this project. All authors contributed equally to this work.

1 Introduction

While the vast majority of countries today hold elections to choose their leaders,[1] only about half of these elections are truly free and fair (Bishop and Hoeffler 2016). Similarly, recent evidence shows that one in three voters worldwide lives in a country where electoral quality has declined, and that the credibility of elections has eroded sharply in recent years (International IDEA 2024). These problems are especially acute in low- and middle-income countries, where elections tend to score lower on quality and are often undermined by irregularities such as fraud, vote buying, and violence (Collier and Vicente 2012; Norris et al. 2014).

These trends are troubling from the perspective of canonical political economy theories, which hold that well-functioning elections improve citizens' welfare by allowing voters to select their preferred leaders (a *selection channel*) and to discipline their effort once in office (an *accountability channel*) (Fearon 1999; Ferejohn 1986; Besley 2006). When elections are compromised by irregularities, both mechanisms may break down: politicians can win office through manipulation rather than by serving voters' interests, and the electoral incentives that ordinarily discipline incumbents lose force. As a result, weaker electoral integrity may lead to underperforming politicians, and ultimately lower welfare for citizens.

These implications carry particular weight in lower-income settings, where access to basic public goods and services, such as sanitation and road infrastructure, remains considerably limited (Gethin 2023), heightening citizens' dependence on the performance of politicians to secure their living standards. The potential returns to improved electoral quality may therefore be highest precisely where elections function most poorly.

Despite these theoretical and practical considerations, systematic evidence on whether electoral integrity improves citizens' lives, and on how much is at stake when it fails, remains scarce. In this paper, we address this gap by leveraging a unique opportunity for experimental variation in electoral integrity. Specifically, we analyze the downstream effects of a randomized intervention implemented during Colombia's 2019 local mayoral elections, designed to test whether a campaign promoting citizen oversight could reduce electoral irregularities, such as fraud, vote buying, and voter intimidation, which are common in this setting (Fergusson et al. 2018; García-Sánchez and Pantoja 2015). The intervention was randomized at the municipal-level across nearly 700 municipalities, representing more than 60% of the country's 1,123 municipalities. It consisted of a large scale social media campaign encouraging citizens to report electoral irregularities through an online channel hosted by an NGO, which then passed these reports to judicial authorities for investigation. In Garbiras-Díaz and Montenegro (2022), we showed that this intervention successfully reduced electoral

[1] According to Varieties of Democracy (V-Dem Dataset), 82% of countries held national elections in 2023.

irregularities, as captured by multiple measures, including mentions in the media, forensic tests, and citizen surveys. We also found that candidates who relied on electoral irregularities experienced a significant decrease in their vote share in treated municipalities.

We leverage the exogenous variation in electoral integrity generated by this experiment to study whether cleaner elections improve local development. We focus on the set of public services that fall directly under the responsibility of local politicians, and thus serve as a transparent, ex-post measure of their performance during their four-year mandate. These include access to sanitation—through clean drinking water, sewerage, and waste disposal—along with public lighting and road infrastructure, services that remain far from universal across the country. For instance, in the municipalities in our sample, access rates to sewerage, water, and waste collection averaged only 42%, 59% and 50%, respectively in 2019.[2]

Measuring the provision of public services presents significant challenges due to the scarcity and often limited reliability of available data. We address these issues by combining three data sources. First, we use administrative data to compute the percentage of properties in each municipality with access to sewerage, aqueduct, and waste collection services, as reported by the respective local service providers. Second, we measure local roads using crowdsourced data from OpenStreetMap, a world-wide repository of data on public infrastructure reported by volunteers. Finally, we use nighttime light intensity as an independent and objective proxy for public lighting, which is otherwise unavailable.

Our main result is that, in the four years following the intervention, treated municipalities experienced an increase of approximately 0.13 standard deviations ($p < 0.01$) in an index of access to the five local public services. This improvement emerged gradually over time and was spread across the entire distribution of municipalities rather than being driven by a few. Examining each service separately, we find uniformly positive effects of similar magnitude in urban and rural areas, indicating that the gains did not come at the cost of tradeoffs across service types or locations.

Effects are typically larger in municipalities assigned to a sub-treatment arm in which the main intervention was combined with a letter sent to the universe of mayoral candidates in the municipality informing them about the citizen reporting campaign. In Garbiras-Díaz and Montenegro (2022), we showed that this sub-treatment arm also experienced a larger decrease in electoral irregularities. Together, these two pieces of evidence suggest that the

[2] Although basic in nature, the public services under the purview of Colombian mayors are essential for local development. A well-established literature shows that investments in basic sanitation infrastructure play an important role in ensuring local populations' health, and in particular in reducing child mortality (for a review see Augsburg et al. 2025). Similarly, road infrastructure, particularly rural local roads, can improve development outcomes and reduce poverty (e.g., Khandker et al. 2009; Asher and Novosad 2020; Aggarwal 2018; 2021). Finally, public lighting has been robustly associated with reductions in crime rates (for a review see Welsh et al. 2024).

increase in service provision is driven by improvements in electoral integrity.

How did these improvements in public services materialize? We turn to the two channels that the political economy literature has emphasized, through which cleaner elections may improve governance: the *selection* of politicians and their *accountability* once in office. We provide evidence consistent with both channels.

To study selection, we examine whether the intervention affected the type of mayor elected. While many candidate characteristics could have changed as a result of the intervention, we focus on the one that should have been impacted most directly: whether a candidate is "clean," in the sense of being unlikely to engage in electoral irregularities. We measure this trait using a survey we conducted prior to the intervention, in which we asked citizens in each municipality to assess candidates' likelihood of engaging in such irregularities. Using this measure, we find that the intervention increased the probability that a clean candidate was elected by 7.1 percentage points ($p < 0.1$). This represents a substantial 23% increase over the control group, in which clean candidates were elected in 31% of elections.

We further characterize the selection channel by examining other candidate characteristics that could be correlated with "cleanliness" and may therefore help explain the results.[3] We compile a comprehensive dataset of politician characteristics—demographics, education, public- and private-sector experience, political experience, political alignment, party affiliation, and campaign funding—and conduct two exercises. First, we test whether the intervention changed these observable characteristics among elected mayors and find no evidence that it did. Second, we examine whether electoral cleanliness is correlated with the same characteristics and find that these correlations are generally weak, with only a few statistically significant but still small exceptions. These results thus suggest that the selection effect on candidate cleanliness is unlikely to be driven by changes in other observable politician traits. However, we cannot rule out correlations with unobservable traits, such as honesty or other psychological characteristics that we are unable to measure.

Next, we investigate the accountability channel. We argue that it could have operated through two routes. First, the intervention may have increased electoral transparency beyond the 2019 election. This channel requires that citizen monitoring of elections persisted over time, giving forward-looking mayors incentives to perform better in anticipation of cleaner future elections in which they, their parties, or their allies might run. Second, the intervention may have increased citizen oversight of politicians through non-electoral means, such as

[3]The relationship between electoral cleanliness and other attributes that could influence public service provision is theoretically ambiguous. Cleaner candidates may be more educated or supported by more competent teams if they rely less on electoral irregularities. Conversely, politicians who engage in electoral irregularities may be better connected to the private sector and may use such irregularities as a signal of these connections to voters (Kramon 2016).

protests or claim-making. We find evidence consistent with the former channel but not the latter. Specifically, in the 2023 local elections—four years after the original intervention—citizens in treated municipalities filed roughly 25% more electoral-irregularity reports than control municipalities. Although smaller than the original short-term effect of over 60%, this evidence indicates that the intervention generated sustained electoral accountability that may have pressured incumbents in treated municipalities to perform better.

In contrast, we find no evidence of increased non-electoral accountability. We examine both collective claims on local governments, measured through protests and participation in grassroots organizations, and individual judicial claims, measured through constitutional writs (*tutelas*). We also examine social media engagement around political and social issues as a broader proxy for citizens' capacity to mobilize. Across measures, we find no evidence that treated municipalities experienced greater citizen action.

We then develop a decomposition that quantifies the contribution of the selection and accountability channels to the overall effect on local public services. Under a set of plausible assumptions, the selection channel can be expressed as the product of two estimates: the change in the probability that a "clean" candidate is elected, and the change in the provision of local public services at the municipal level given that a clean candidate is elected.

The first term is the 7.1-percentage-point increase in the probability of electing clean candidates documented above. The second term comes from a close-election regression discontinuity design comparing access to services in municipalities where a "clean" candidate barely won to those where a "dirty" candidate barely won. This design isolates races in which candidate type is as good as randomly assigned, holding the accountability channel fixed.[4] We estimate that electing a clean candidate increases the public service index by 0.41 standard deviations (p-value< 0.05).[5] Combining the two terms, we find that approximately 22–30% of the estimated average treatment effect is attributable to the selection channel. Under the additional assumption that channels unrelated to the mayor are shut down, the remaining 70–78% can be attributed to the accountability channel.

Having documented evidence consistent with both selection and accountability, we next examine the concrete actions through which mayors could have improved public service provision. We focus first on public procurement—a critical stage in service delivery in which

[4] While this design identifies a local average treatment effect of electing a clean candidate in close elections, we argue that this is precisely the object of interest for the decomposition: the selection channel should only materialize in marginal races in which the candidate type elected "flips" as a result of the intervention.

[5] As noted by Marshall (2024), regression discontinuity designs that study the effect of a particular politician characteristic will likely confound the effects of other associated characteristics, especially in the presence of compensating differentials. In our setting, however, we find that other observable characteristics of mayors do not change discontinuously at the cutoff, echoing the results mentioned above that cleanliness does not seem to capture other observable characteristics of mayors.

corruption and inefficiency frequently arise in the allocation of government resources, both in Colombia (Gulzar et al. 2022) and in other developing contexts (Bosio et al. 2022). We find that treated municipalities rely more on competitive procurement, such as public tenders and auctions, and less on non-competitive forms, such as direct contracting, which prior work links to more politically connected firms and lower efficiency (Fazekas et al. 2025; Baltrunaite et al. 2020). We then examine whether the intervention affected corruption by local government officials, which has been found to be an important source of distortions for local economies (Colonnelli and Prem 2022; Olken 2007; Ferraz et al. 2012). We proxy for it using corruption-related judicial processes, and find that treated municipalities face fewer such processes, pointing to an additional way in which mayors appear to have improved the use of public funds. Together, these findings indicate that stronger electoral integrity can lead to better governance by encouraging politicians to adopt more transparent and efficient mechanisms for allocating resources.

Finally, we rule out two alternative mechanisms. First, we show that treated municipalities do not collect more revenue, either in total or from specific sources such as local taxes or national transfers. This suggests that the improvements in public services are more consistent with efficiency gains rather than with increased funding. Second, we show that the effects are not driven by additional support from higher levels of government. To do so, we examine services that fall outside the mayor's mandate—namely, education, healthcare, and national roads—and find no statistically significant differences across treatment arms.

Related literature: This paper relates most closely to a small but growing literature on the relationship between electoral irregularities and economic outcomes. Many studies in this literature provide correlational evidence, while fewer offer causally identified estimates.[6] Guerra and Justesen (2022) use a lab experiment in Kenya to show that vote-buying reduces the demand for public goods. Closer to our paper, Ofosu (2019) studies a field experiment in Ghana that randomized electoral observers in parliamentary elections. He finds that legislators elected in more heavily monitored constituencies spent more development funds, which he interprets as increased politician responsiveness. Similarly, Karim (2025) studies a voter re-registration reform in Brazil aimed at curbing vote-buying. Using a regression discontinuity design, he finds that the reform increased political competition and reduced the incumbency advantage—a result consistent with Hidalgo and Nichter (2016). In the years following the reform, treated municipalities experienced higher expenditure on health

[6]For instance, Khemani (2015) documents a negative association between vote-buying practices and public health outcomes in the Philippines, while Anderson et al. (2015) show that clientelistic practices are associated with fewer pro-poor policies.

and education, as well as improvements in related outcomes such as infant mortality.[7]

We contribute to this literature in four ways. First, to our knowledge, we provide the first experimental evidence that improving electoral integrity increases downstream access to local public services, outcomes inherently connected to citizen well-being. Second, we move beyond reduced-form evidence by documenting the mechanisms through which electoral integrity improves politician performance: we find evidence of both selection and accountability channels, as well as specific intermediate actions by politicians that underpin the improvements in local public services. Third, we develop a novel framework to quantify the relative importance of selection and accountability in explaining the overall effect of the intervention. Finally, rather than relying on selected measures of performance, we construct a comprehensive index based on the set of public services assigned to local governments in our setting, and test whether effects vary across services and across rural and urban areas within municipalities. This is important, since politicians could improve access to only certain types of public services or geographic areas at the expense of others.

In addition to these studies on the *quality* of elections,[8] our paper also speaks to a more established strand of work on the economic effects of democracy at the "extensive" margin. While earlier studies tended to find mixed results, more recent ones find stronger and more consistent evidence linking democracy and democratization events to economic growth as well as to human capital investments.[9] Our findings complement this literature by showing that, even after democratic institutions are in place, the quality of elections is a crucial determinant of whether democracy fosters development and improves citizens' welfare.

A second strand of related literature examines the selection and accountability channels, which are the main proposed mechanisms through which democracy promotes citizen well-being (Fearon 1999; Ferejohn 1986; Besley 2006). The selection channel has typically been tested by studying whether information about candidates' past performance or qualifications affects voters' choices (e.g., Ferraz and Finan 2008; Dunning et al. 2019; Casey et al. 2021), by understanding who runs for office and why (e.g., Dal Bó et al. 2017; Dal Bó and Finan

[7] Relatedly, Blattman et al. (2026) study a randomized intervention to reduce vote-buying in Uganda. While the intervention failed to reduce vote-buying, it increased electoral competition faced by incumbents. Furthermore, student test scores in schools close to the treated parishes increased one year after the intervention, which the authors interpret as increased incumbent effort in response to competition from challengers.

[8] Other dimensions of electoral quality studied in this literature include political competition (see Gottlieb and Kosec 2019, for a review), improved voting technology (Fujiwara 2015), the introduction of local elections to better target local populations' preferences (Martinez-Bravo et al. 2022; Paniagua 2022), and the effect of political turnover on economic performance (Marx et al. 2024; Akhtari et al. 2022; Bazzi et al. 2025).

[9] Starting with the seminal work by Przeworski (2000), these relationships have been studied extensively. For a review, see Doucouliagos and Ulubaşoğlu (2008) for early contributions to this literature, and Colagrossi et al. (2020) for more recent work. In one of the most recent and meticulous studies in this literature, Acemoglu et al. (2019) show that one of the mechanisms through which democracy fosters growth is by increasing public goods provision—a result that aligns with our findings.

2018; Bowles and Larreguy 2025), and by examining whether candidate characteristics—such as gender or pro-sociality—influence their policies (e.g., Chattopadhyay and Duflo 2004; Beath et al. 2016; Gulzar and Khan 2024). Similarly, several papers test the accountability channel by exploiting variation in term limits to assess how incentives affect politician performance (e.g., Ferraz and Finan 2011; Coviello and Gagliarducci 2017). Yet a third set of papers has sought to disentangle both channels, either using specific quasi-experiments (Alt et al. 2011; Gagliarducci and Nannicini 2013) or structural models (Sieg and Yoon 2017; Aruoba et al. 2019). Our paper contributes to this literature by introducing a simple method to quantitatively decompose the selection and accountability channels, which can be readily used in other settings. In doing so, we also engage with recent work clarifying the conditions under which politician regression discontinuity designs identify the effects of candidate characteristics (Marshall 2024), and we discuss how these conditions apply to our strategy.

Finally, this paper contributes to the broader literature on the determinants of local public goods provision. A large body of research has emphasized the effect of public sector employees and civil servants on the efficiency of service delivery, focusing particularly on the impact of incentives, technological innovations, and bureaucratic quality (e.g., Carreri 2021; Muralidharan et al. 2021); for a review, see Finan et al. (2017). Related work highlights the roles of party networks, citizen mobilization, and ethnic divisions in shaping access to local services (Auerbach 2020; Björkman and Svensson 2009; Habyarimana et al. 2009). Our findings complement this literature by underscoring the importance of electoral quality as a critical upstream driver of service provision and more efficient government spending.

The remainder of the paper is organized as follows. Section 2 describes the context, including the 2019 intervention and the institutions regulating local politicians' responsibilities in Colombia. Section 3 outlines the data sources, the construction of variables, and the empirical strategy. Section 4 presents the main results on public service provision. Section 5 examines mechanisms: 5.1 presents our methodology to decompose the main effect into selection and accountability channels, 5.2 studies mayors' intermediate actions, and rules out alternative explanations. Section 6 concludes.

2 The Colombian Context

Colombia offers an ideal setting to study the relationship between electoral integrity and local development. Although the country has maintained formal democratic institutions for most of its history—interrupted only by a brief period of military rule between 1953 and

1957—elections continue to be marred by widespread irregularities (MOE 2023). Clientelism, in particular, remains pervasive: recent surveys indicate that about 18% of respondents have "sold" their vote at least once, and around 7% report coercion or intimidation to vote in a particular way (Fergusson et al. 2018; García-Sánchez and Pantoja 2015). These rates are substantially higher among rural and lower-income populations.

Against this backdrop, online reporting of electoral irregularities has become an increasingly important oversight mechanism. The *Misión de Observación Electoral* (MOE), a prominent domestic NGO specializing in election monitoring, operates the country's most widely used online platform, which allows citizens to *anonymously* report instances of vote buying, coercion, and other forms of electoral misdeeds. These reports are forwarded to the relevant national oversight bodies—most notably the Attorney General's Office and the Inspector General's Office—for formal investigation. This collaboration between the MOE and national watchdogs makes online reporting a consequential instrument of accountability: between 2010 and 2018, approximately 15% of reports led to formal investigations and 2.5% resulted in judicial decisions such as sanctions or acquittals (Garbiras-Díaz and Montenegro 2022).[10]

2.1 Mayoral Elections

Mayors in Colombia are elected to four-year, non-renewable terms by simple plurality.[11] Elections are held concurrently with those for governors, municipal councils, and departmental assemblies. Candidates may run under a party label or as independents, supported by citizen signatures.

Although immediate reelection is prohibited, several mechanisms ensure that mayors remain responsive to their constituents and accountable for their performance in office. First, formal oversight institutions—including the Office of the Inspector General (*Procuraduría General de la Nación*), the Office of the Comptroller General (*Contraloría General de la República*), and the Office of the Attorney General (*Fiscalía General de la Nación*)—regularly audit and investigate potential cases of corruption, embezzlement of public funds, irregularities in procurement, or any type of mismanagement by municipal governments, and can impose sanctions when necessary. Civil society organizations may also initiate recall processes or escalate concerns to national oversight bodies.

Second, career incentives sustain performance despite term limits. Local office frequently

[10] Even when investigations do not lead to formal penalties, they often impose substantial reputational costs on politicians, as the proceedings and their outcomes are publicly accessible.

[11] An exception applies in Bogotá, where a runoff is held if no candidate obtains at least 40% of the vote and the leading candidate does not win by a margin of at least 10 percentage points.

serves as a stepping stone to future political positions. For instance, analyzing the universe of candidates running for local and legislative office between 1997 and 2023 reveals that over 40% of elected mayors run again in a later election—local or legislative—and around 30% rerun in their own municipality after the mandatory cooling-off period (see Figure A1).[12] These patterns indicate strong electoral ambition among mayors, providing a reason to exert effort in office even without the prospect of immediate reelection.

These career incentives extend beyond future electoral candidacies. Many mayors belong to parties or local family political dynasties that depend on well-performing representatives to maintain influence. These actors can reward or punish performance through internal promotions, endorsements for higher-level positions, or the withdrawal of future opportunities (Caputo-Leyva 2022). Finally, even in the absence of party or dynasty pressures, mayors frequently pursue political or professional careers that hinge on their reputation as competent and effective leaders, much of which is shaped during their term in office.

2.2 Local Service Provision and Mayoral Responsibilities

Mayors serve as the highest executive authority at the municipal level, responsible for implementing the local budget, ensuring the provision of basic public services, and coordinating with national and departmental authorities. As in other Latin American countries, these responsibilities have been gradually strengthened through a process of decentralization over the past decades, particularly following the 1991 Constitution (Falleti 2010).

The Constitution, together with key pieces of legislation, assigns mayors a well-defined set of service-delivery responsibilities that are not shared with other levels of government. This institutional arrangement makes it easier to link outcomes in these domains to municipal governments. In the next section, we describe how we measure performance across several dimensions of service provision.

Mayors are primarily responsible for five public services: (i) aqueduct and (ii) sewer network connectivity, (iii) waste collection, (iv) public lighting, and (v) local road infrastructure. The first three of these services are known as "residential public services" (*servicios públicos domiciliarios*), and are regulated by *Law 142 of 1994*. The latter two are classified as "non-residential services" and are regulated, mainly, by *Decree 2424 of 2006* and *Law 136 of 1994*. While road infrastructure and public lighting that connect more than one municipality may be provided or supported by higher-level governments, it is the explicit responsibility of

[12] We use the administrative records compiled by CEDE at the University of Los Andes. We match elected mayors to candidates in subsequent elections using a phonetic algorithm based on full names and two last names. To reduce false positives, we restrict to individuals with two last names and require exact municipality matches for reruns in the same locality. These choices reduce misidentification risk but may omit some true matches.

mayors to ensure the provision of these services when they fall entirely within municipal boundaries.

Mayors are tasked with ensuring that these services are provided broadly, efficiently, continuously, and with adequate quality. The tools they are endowed with to do so are several. First, they are in charge of formulating development plans (*Planes de Desarrollo Municipal* and *Planes de Ordenamiento Territorial*) to set service provision goals and allocate the budget intended to cover these expenditures. Municipal revenues come from a combination of locally collected taxes, fees, and service charges, as well as transfers from the national government. A substantial share of these transfers is earmarked for specific purposes, including the provision of basic public services.

Second, they decide *how* these services are provided. They may do so directly through publicly constituted companies—particularly in small or underserved areas—or by contracting their provision to private or mixed companies through procurement processes.[13] Procurement is a central component of the production function of public services and remains vulnerable to various types of inefficiencies. In this context, as in other under-institutionalized settings, it has been shown to be affected by rent-seeking, collusion, and political favoritism, especially at the subnational level (Gulzar et al. 2022). Thus, mayoral supervision and involvement in this process play an important role in improving service delivery and quality.

Third, mayors are responsible for monitoring the provision of these services by local providers. Although they do not hold direct sanctioning authority, they can escalate concerns to national-level entities—most importantly, the *Superintendencia de Servicios Públicos Domiciliarios*—which is empowered to investigate providers, impose sanctions, and, in extreme cases, intervene in their operations. This oversight role allows mayors to influence service quality indirectly by ensuring that service failures and irregularities are reported to the relevant regulatory authority.

Access to locally provided public services is limited and uneven throughout the country. In 2019, just before our intervention, median property access rates for aqueduct, sewerage and waste collection were 57.5%, 37% and 42.4%, respectively. Moreover, access varies widely across municipalities—for example, aqueduct coverage ranges from 96.8% (90th percentile) to 27.7% (10th percentile)—and within municipalities, where rural access often represents only 18%-40% of urban access. While gaps are partially mitigated by non-conventional or individualized solutions (e.g., latrines, rainwater collection), disparities remain substantial.

[13] Municipalities may provide services directly only when no private providers are willing to do so, or when they can demonstrate that they can deliver the service more cost-effectively than private providers. As a result, direct public provision is relatively rare; in 2023 only about 9% of providers were public (Superservicios 2024).

Other public services. Contrary to other countries in the region, such as Brazil, Colombian mayors are typically not responsible for health or education services.[14] In these sectors, mayors play primarily a coordinating role, while responsibility and financing largely rest with departmental and national authorities. Finally, although mayors are formally designated as the highest security authority at the local level, their role in practice is limited. Security enforcement is carried out mainly by the National Police and the military. Mayors coordinate activities with these institutions and may contribute to initiatives such as surveillance systems; however, priorities and resources for personnel, equipment, and operations are set and managed by the national government.

2.3 The 2019 Experiment

In our original intervention, we implemented a large-scale field experiment around the 2019 local elections to address two common shortcomings of conventional election-monitoring strategies: limited scalability and low cost-effectiveness. To do so, we tested whether a citizen-based election monitoring campaign—disseminated online and implemented in coordination with enforcement authorities—could serve as a scalable and cost-efficient tool for deterring electoral irregularities. While the full details of the intervention and its results are presented in Garbiras-Díaz and Montenegro (2022), we briefly summarize here the main features of the experimental design and its short-term effects on irregularities and electoral outcomes.

Experimental design. The unit of randomization of the experiment was the municipality, the administrative level at which mayors are elected. The sample comprised 698 municipalities.[15]

A randomly selected two-thirds of these municipalities were assigned to the main treatment, which was designed to encourage citizens to report electoral irregularities through an online platform called "*Pilas con el voto*" ("Watch out for the vote"), operated by the MOE (see above for a description of this NGO and its operations). To do so, we implemented a large-scale social media campaign delivered via Facebook and Instagram ads. The campaign ran for five days—from October 24 to October 28, just before and during the October 27

[14] Departmental capitals and some municipalities with populations above 100,000 that demonstrate high financial and administrative capacity can be certified to provide these services (*Law 715 of 2001*). Given that the selection criteria for municipalities in our intervention had a relatively low population threshold, very few of them are certified in our sample.

[15] This sample was selected based on population criteria, targeting municipalities between 5,000 and 97,000 inhabitants. This restriction was meant to guarantee that a substantial share of the population saw the ads given a limited budget.

elections—and targeted all adults (18 years and older) residing in treated municipalities.[16] The campaign reached 4.4 million users—approximately one-third of the targeted population.

The main ad shown in treated municipalities included two components: (i) an information message indicating citizens where to report, thereby lowering the cost of reporting; and (ii) a call-to-action message emphasizing the urgency of reducing irregularities.[17] To net out any potential effects of exposure to political ads, control group municipalities received a "placebo" ad reminding citizens about the election date.

In a second treatment arm, we coordinated with the Attorney General's Office to send formal letters to the universe of mayoral candidates in a random subset of municipalities treated with the online campaign. A first letter was mailed on September 6, after the candidate registration deadline, so that the intervention did not affect candidate selection into the elections, followed by a reminder email three weeks before the election. This component was designed to test whether candidates responded to the prospect of increased monitoring.[18]

Results on irregularities. Because electoral manipulation is inherently covert, irregularities are difficult to observe directly. To overcome this challenge, in Garbiras-Díaz and Montenegro (2022), we relied on two complementary approaches to measure them. First, we used mentions of electoral irregularities in local and national media—including radio, television, print, and online outlets—collected by an independent media-monitoring firm. Second, we constructed a forensic measure using fine-grained vote-tally data at the voting booth level (*mesa de votación*), applying standard tests from the criminology literature, such as second-digit Benford's Law, to detect statistical anomalies consistent with manipulation.

Across both measures, we find that the social media intervention reduced the incidence of electoral irregularities. Treated municipalities were 6 percentage points less likely to be mentioned in media reports documenting electoral misconduct. Using our forensic measure, treated municipalities were 9 percentage points less likely to fail tests for statistical anomalies.

[16] Electoral crimes are defined in the Colombian Criminal Code. Misdeeds occurring before election day include campaigning by public officials and voter registration fraud. Election-day violations include electoral fraud and illicit political advertising, which is prohibited on election day. A full list is available at https://www.cne.gov.co/delitos-electorales.

[17] In practice we randomized whether municipalities received both components or only one. For conciseness, and to maximize power, we pool these treatment arms in the main analyses.

[18] Among municipalities receiving the letter, we also randomized the version of the letter to account for the possibility of signal jamming—for instance, candidates might use the information about the reporting site to systematically attack the opposition through false reports, thereby rendering the platform uninformative. To that end, we created a "full knowledge" version of the letter that explicitly informed candidates about the website used in the citizen campaign. We did not find substantial differences across these versions of the letter in Garbiras-Díaz and Montenegro (2022). Accordingly, we pool these subtreatments and focus on the overall campaign and the letter arm in our main analyses.

We obtain qualitatively similar results using continuous versions of both measures.[19]

For the second treatment arm—the letters sent to candidates—most effects appear to be driven by municipalities where candidates were explicitly informed about the campaign, suggesting a potential deterrence mechanism behind the observed reduction in irregularities. However, the estimated difference in coefficients is not statistically significant.

Results on electoral outcomes. We next examined whether the intervention affected electoral outcomes. Our primary test analyzed whether candidates who relied more heavily on electoral malpractice lost vote share in treated municipalities. To identify these candidates, we administered a pre-treatment survey three weeks before the intervention, asking citizens to evaluate each mayoral candidate's perceived propensity for electoral misconduct. We find that candidates scoring above average on this irregularity measure experienced a 2.5 percentage point reduction in vote share in treated municipalities. Beyond this penalty, we find no evidence that the intervention affected other electoral outcomes, such as turnout or political competition.

3 Data, Outcomes, and Empirical Specification

3.1 Measuring Local Public Service Access

Analyzing locally provided public services presents several challenges. Data are often incomplete, inconsistently collected, or subject to quality concerns. We address these challenges by combining different sources depending on the type of public service studied.

Administrative data. We collect data on municipal access to aqueduct, sewerage and waste collection from the *Superintendencia de Servicios Públicos Domiciliarios* (SSPD), a national public agency overseeing the provision of these "domiciliary" public services. By law, municipalities must report these data annually, with the help of providers, to a unified system managed by the SSPD (Law 1176 of 2007). The main indicators measure the share of properties in a municipality with access to each of the mentioned public services.

We implement two data-cleaning steps on the 2016-2024 available data. First, we interpolate to fill a few missing values present in the data. Second, we clean for outliers by replacing observations at time t with the average of the values at $t-1$ and $t+1$, when the value at t differs by more than three standard deviations from this average. This step

[19] The original article also presents results for additional outcomes and a series of robustness checks.

corrects occasional implausible breaks in the time series that are occasionally observed, and which are likely due to the fact that the data are self-reported.

Open-source data. The Colombian government does not systematically collect data on "tertiary" road infrastructure, which comprises the roads under municipal jurisdiction. To overcome this limitation, we utilize OpenStreetMap (OSM), an open-source platform where a global community of volunteers compiles GIS data on road networks. OSM has been previously employed to measure infrastructure in low-income contexts typically characterized by data scarcity (e.g., Dumas and Játiva (2025) and Barrington-Leigh and Millard-Ball (2017)). Crucially, each road in the repository is categorized by type, allowing us to distinguish larger, national highways from local roads.[20]

We measure the total kilometers of local roads within the boundaries of each municipality across different OSM data snapshots. As with any crowdsourced dataset, reporting may lag construction or include planned infrastructure, making the exact timing of road availability difficult to determine. Despite these limitations, OSM provides a highly valuable proxy for local road infrastructure in this data-scarce environment.

We use a $\log(x + 1)$ transformation of this variable to adjust for its right-skewed distribution, as well as to account for the fact that a few municipalities do not have any reported roads at baseline, in 2019.

Satellite data. Data on public lighting are not systematically collected by the central government either in our setting. To overcome this lack of information, we use a historic satellite-based proxy for this public good. In particular, we use annual nighttime light data from the *Visible Infrared Imaging Radiometer Suite* (VIIRS). Nighttime lights have been widely used in political science and economics for a variety of purposes, including to proxy for electrification efforts in developing contexts (see Gibson et al. (2020) for a review). The public version of these data aggregates cloud-free monthly radiance across pixels measuring approximately 500x500 meters. We use the median radiance across months to minimize the risk that outliers introduce noise into the estimates. As before, we apply a log transformation to account for the variable's right-skewed distribution.

Survey data. The National Department of Statistics (DANE, by its acronym in Spanish) collects household-level information on access to aqueduct, sewerage, and waste collection services through the yearly Quality of Life Survey (*Encuesta de Calidad de Vida*, or ECV).

[20] In practice, we define local roads as those not tagged as "primary," "secondary," "trunk," or "motorway," as these classifications strictly fall outside municipal jurisdiction.

While this survey is not representative at the municipal level, it provides an external validation of the administrative data, which are self-reported and may raise concerns about data quality. We therefore use these survey responses as a robustness check.

We accessed the microdata through DANE's secure access facilities and pooled the ECV datasets for the years 2022 and 2023 to maximize coverage. The resulting sample includes approximately 103,000 respondents across 469 municipalities. We conduct the analysis at the household level, allowing us to include an indicator for whether the residence is located in a rural or urban area as a control.

3.1.1 Public Service Index

We aggregate the administrative and satellite-based indicators described above into a single index. This composite measure provides a parsimonious way to assess whether the intervention affected public service provision overall. It also helps mitigate concerns related to multiple hypothesis testing given the relatively large number of individual outcomes analyzed (Anderson 2008). As a complement, we report results for each service separately.

We construct our main index in three steps. First, we measure changes in each public-service outcome:

$$\Delta y_m^j = y_{m,post}^j - y_{m,pre}^j$$

where $y_{m,post}^j$ is the measure of public good of type $j \in \{\text{aqueduct, sewerage, waste, roads, lighting}\}$ in municipality m, measured in a period after the intervention, while $y_{m,pre}^j$ is measured before the intervention. In our main analysis, the pre-intervention period is defined as the average across the years 2018 and 2019, while the post-intervention period corresponds to the average across 2022 and 2023. We focus on outcomes toward the end of the mayoral term because expanding the stock of public goods typically takes time. In the results section, we also analyze the dynamics of public goods provision in each year, which confirms this intuition.

Our main analysis uses the *changes* in public good provision in the construction of the public service index, as described above, rather than the *levels*. This approach improves the precision of the estimates by accounting for pre-existing municipal-level differences in public goods. As a robustness check, we also report results based on the levels of public goods provision and provide evidence that there are no significant imbalances in these variables during the pre-intervention period.

Second, we standardize each of these variables using the mean and standard deviation of

the municipalities of the control group.

$$z\left(\Delta y_m^j\right) = \frac{\Delta y_m^j - \mu_{\Delta y_m^j, control}}{\sigma_{\Delta y_m^j, control}}$$

Finally, we average these standardized variables to construct our final index:[21]

$$I = \frac{1}{5}\sum_j z\left(\Delta y_m^j\right)$$

3.2 Additional Variables on Mechanisms

We collected data on four groups of variables that shed light on the mechanisms underpinning the effects on public services.

Candidate characteristics (selection). We compile a comprehensive set of characteristics for mayors elected in 2019 to investigate how the intervention may have influenced candidate selection. Candidate age, sex, and party affiliation are obtained directly from the information provided upon registering their candidacy with the *Registraduría Nacional*, the primary electoral authority. To estimate political experience, we count the number of times each candidate has been elected to municipal councils, mayors' offices, the Senate, or the House of Representatives since 1992, using election data collected and cleaned by the *Centro de Estudios sobre Desarrollo Económico* (CEDE) at the *Universidad de los Andes* and updated via a freedom of information request to the Registraduría Nacional to include the 2022 and 2023 elections, which Los Andes had not yet processed. Additionally, we gather information on mayors' education and their experience in the public and private sectors from the biographical statements included in their candidacy manifestos, which are submitted to and publicly posted by the *Registraduría Nacional*.[22] We also include campaign finance information scraped from the *Consejo Nacional Electoral*'s *Cuentas Claras* portal.[23] We use these data to measure the share of campaign income from private donations, self-funding, and party contributions, as well as total reported income (logged). Finally, we collect data

[21] Three municipalities in our sample lack administrative data on aqueduct, sewerage, and waste collection coverage. For these municipalities, the index is computed as the average of the road and nighttime-light indicators.

[22] Although submission of manifestos is legally required, only 462 mayors—approximately two thirds of our experimental sample—include biographical information. In our main analysis, we treat missing bios as missing data; however, results remain robust when coding these cases as zeros (i.e., indicating no education or public/private sector experience). Importantly, the submission of bios is well balanced across treatment arms.

[23] Specifically, we use *Formulario 5B* reports.

on mayors' business ownership by scraping records from the *Registro Mercantil*, Colombia's official business registry.

Non-electoral accountability. We consider four types of outcomes that capture the primary ways in which citizens in this context may exert non-electoral accountability, by either expressing their desire for improved public goods or organizing collectively to provide them directly. The first is the incidence of citizen protests during the study period (2020–2023), measured using the *Base de luchas sociales* compiled by the *Centro de Investigación y Educación Popular* (CINEP). This database is constructed through manual coding of protest events reported in newspapers, radio, television, online media, social media, and organizational communications.

The second variable is the number of, and membership in, *Juntas de Acción Comunal* (JACs), which are legally regulated grassroots organizations that operate at the local level and play a key role in civic participation and local development (Law 743 of 2002). These organizations serve as intermediaries between citizens and local governments, advocating for community needs and priorities. They also occasionally implement small-scale infrastructure projects, particularly in areas such as sanitation and local road construction. With over six million members nationwide, JACs constitute the largest grassroots organization in Colombia dedicated to local development. We obtained data on the number of JACs and their membership per organization through a freedom of information request sent to the Ministry of the Interior, the institution responsible for regulating their operations.

The third variable is the number of submitted *tutelas*, which are constitutional writs filed by individuals seeking protection of their fundamental rights before the state. These often include demands for access to basic public goods, given that such access is constitutionally protected in Colombia. The *tutela* process is designed to be accessible: it does not require legal representation, and judges are required to issue an initial ruling within ten business days (Law 393 of 1997). As a result, tutelas have become a widely used mechanism for citizens to defend their constitutional rights (see Londoño-Vélez and Saravia (2025) for further context). We use publicly available aggregate data from the Constitutional Court on the number of *tutelas* filed in each municipality between 2020 and 2023, along with information on the specific rights invoked in each case.

The final set of variables related to non-electoral citizen accountability is social media engagement. We collect these data using Meta's Marketing API, which provides lower and upper bounds on the number of monthly active users in a targeted population. Our measure is the average of these two bounds. We measure both the total number of users aged 18 and older in each municipality and the subset interested in political and social issues, according to

Meta's classification.[24] Unfortunately, this data is not available retrospectively. We collected it in March 2026, and it is thus only a proxy of the number of users in our study period.

Mayors' governance actions. We gathered information on two governance-related actions that mayors might have taken to improve access to local public services. The first is related to procurement practices, and comes from SECOP, a publicly available dataset that records all public procurement contracts executed by all levels of government in Colombia.[25] We restrict the sample to contracts executed by municipal governments and councils, excluding autonomous public entities such as hospitals, utilities, and educational institutions.

This dataset includes information on the contracting government agency, the final provider of the good or service, the total value of each contract, and a description of the contract's purpose. Critically, SECOP also reports the modality of the procurement process—that is, whether the contract was awarded through a competitive selection process or assigned directly to a provider without competition. We use this variable to calculate, for each municipality, the share of total contract value awarded competitively versus directly over the 2020–2023 period. We define competitive procurement procedures as those awarded through public tenders (*licitaciones*), merit contests (*concurso de méritos*), or auctions (*subastas*).[26]

We collected data on judicial corruption cases involving local governments from the *Sistema Penal Oral Acusatorio* (SPOA) dataset, Colombia's national criminal case management system, which is maintained by the Office of the Attorney General (*Fiscalía General de la Nación*). The dataset records all criminal investigations and allows cases to be linked to the type of government entity involved. We restrict attention to municipal governments and focus on investigations related to corruption offenses.[27] When considering the intensive margin of this variable, we use a $\log(x + 1)$ transformation of this variable to account for outliers.

[24] Specifically, we define these users as those expressing interest in "Politics and social issues", "Politics," "Community issues," or "Volunteering."

[25] For completeness, we combine data from SECOP I and SECOP II, which are semi- and fully-electronic platforms through which governments report procurement.

[26] The information in SECOP is entered by the procurement offices of government entities, so reporting errors and typos are common. To reduce the influence of extreme outliers, we exclude contracts whose ratio of contract value to average annual municipal income exceeds the 99.99^{th} percentile in their category of contracted goods and services. These categories include, for instance, consulting, service contracts, or public works, which have very different typical contract values, warranting doing a category-specific outlier-removal process. We manually confirm that each of the dropped contracts seemed to have a typo in its reported value.

[27] We consider corruption cases to be those labeled as having to do with corruption crimes, which can be further linked to administrative, judicial, tax, electoral or private corruption.

Alternative mechanisms. Finally, we collect data on municipal fiscal capacity and other public services provided by higher levels of government to rule out two alternative explanations for the overall effects.

1. Local government revenue. We obtain data for municipal revenues from the *Departamento Nacional de Planeación* (DNP), which compiles detailed annual information on local government finances reported through standardized administrative forms. Using the official categorization of revenue sources, we analyze differences across municipalities in the composition of revenues, including national government transfers, locally collected taxes, and other income sources.

2. Higher-order public services. We gather data on education and healthcare inputs, which are provided by higher-order government administrations, as explained previously. For the former, we rely on data collected by the Colombian National Statistics Office in coordination with the Ministry of Education through the *Censo Anual de Educación Formal*. We use information on the number of facilities, teachers, administrative staff and students in each municipality, focusing on primary and secondary education. We compute pre- and post-intervention changes using data from 2019 and 2022 (the most recent available wave), respectively. For healthcare inputs, we rely on data from the Colombian Ministry of Health. We draw on multiple sources, including the *Registro Especial de Prestadores de Servicios de Salud* and the *Registro de Talento Humano en Salud*. We construct indicators for the number of facilities, physicians, nurses, beds, offices, and other inputs. We compute outcomes averaging the 2022 and 2023 data, and use the average between 2018 and 2019 as a baseline control.

3.3 Empirical Specification

We use a simple linear specification to estimate the average effects of the intervention:

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

where y_m is the outcome variable in municipality m; $\mathbf{T}_m$ is a vector of indicators for different aggregations of treatment arms in the intervention; X_m is a set of municipal-level covariates; and ϵ_m is the error term.[28]

[28] While our strategy to examine the effects of electoral integrity on public service provision can be understood as through the lens of an instrumental variable framework, in which the effects of the experimental intervention on electoral irregularities represent a "first stage" relationship, we refrain from actually computing IV estimates for two main reasons. First, irregularities are measured with substantial noise given the difficulty in measuring them. As a result, using our intervention as an instrument for irregularities would

In our main specifications, we estimate this equation using two different aggregations of the treatment arms. First, we pool all treatment arms into a single combined treatment indicator. Second, we distinguish between treated municipalities in which we sent letters to mayoral candidates and those in which we did not. We focus on these partitions because, as reported in Garbiras-Díaz and Montenegro (2022), they capture the main differences in electoral irregularities—municipalities where candidates were sent a letter experienced the largest reductions in such irregularities, as discussed above. Nevertheless, we also present results for the "long model", encompassing all combinations of treatment arms, as a robustness check.

Following Garbiras-Díaz and Montenegro (2022), we select the covariates included in X_m using the double-post-Lasso covariate selection method proposed by Chernozhukov et al. (2015) and Belloni et al. (2014). The set of potential covariates is the same as in the original paper.[29] When we use outcomes measured for particular years, rather than the differences used for our public service index, we additionally include a baseline measure of the outcome variable when available. The baseline year(s) depend on data availability, which we specify in the notes to the relevant figures and tables. We also report our main estimates without control variables as a robustness check.

Finally, all tables include randomization inference p-values (Athey and Imbens 2017; Young 2018) alongside heteroskedasticity-robust standard errors.

3.4 Descriptive Statistics and Balance

Table A1 reports summary statistics for the public service index, described in Section 3.1.1, and its five components across the experimental sample. Panel A reports changes between the pre-intervention (2018–2019) and post-intervention (2022–2023) periods, while Panels B–D report the raw (non-standardized) components in changes (Panel B) and in pre- and post-period levels (Panels C and D).

In the post-intervention period, the average municipality had household coverage rates of 64% for aqueduct services, 46% for sewerage, and 53% for waste collection, with substantial

yield a first stage that is severely attenuated by measurement error. This would generate weak-instrument problems—such as highly unstable and potentially inflated IV estimates—and thus provide little information about the local average treatment effect (Staiger and Stock 1997). Second, the exclusion restriction might potentially fail to hold (which we study in the alternative explanation section), further complicating the interpretation of the IV estimate. For these reasons, we interpret our estimates as reduced-form effects of cleaner elections rather than the causal effect of irregularities per se.

[29] These include measures of reporting to the MOE, log population, population density, rurality rates, poverty rates, per capita income, homicide rates, Facebook penetration rates, region dummies, and a large set of political covariates, such as past turnout and election outcomes for different parties, as well as past values of the forensic tests.

variation across municipalities.[30] Average coverage increased by roughly 3–5 percentage points relative to the pre-intervention period. Local roads (measured as log kilometers) and nighttime lights (log mean luminosity) also increased on average, by 0.71 and 0.34 log points, respectively—or ~100% and ~40% relative increases.

Table A2 reports summary statistics for the mechanisms we explore in Section 5.2: procurement practices and corruption-related judicial investigations against the mayor's office. During the post-intervention period, municipalities allocated on average 45% of the total value of their contracts through competitive procedures and 41% through direct procurement, with substantial heterogeneity across municipalities. The average municipality experienced 1.5 corruption-related investigations, while approximately half of municipalities had at least one investigation. Relative to the pre-intervention period, competitive procurement declined modestly, direct contracting increased, and corruption-related investigations became slightly more common.

Figure A2 reports balance checks for each of these outcomes (public service index, components, and the full set of mechanism variables). Concretely, it reports the coefficients of regressions on the pre-period values on indicators for the pooled treatment arm (panels on the left) and for the Letter and No-Letter sub-arms (panels on the right). We report coefficients in control-group standard-deviation units so magnitudes and confidence intervals are comparable across heterogeneous outcomes. Across these variables the estimates are statistically indistinguishable from zero at conventional levels, consistent with successful randomization. We also explore balance for the alternative mechanisms we explore in Section 5.2.1. Overall, out of the around 50 balance tests reported in Figure A2, the number of marginally significant deviations is well within what we would expect under the null.

4 Main Results: Electoral Integrity and Public Service Provision

We begin by examining the effect of the intervention on local public service provision during the four years following the online campaign. As discussed in Section 2, in our setting, mayors are primarily responsible for providing these services, and thus progress made on this front represents a natural measure of mayors' performance in office.

Estimates from equation (1), reported in Figure 1 (and Table A3), show that treated municipalities exhibit better service provision four years later, relative to their pre-intervention

[30] Three municipalities in our sample exhibit waste collection coverage slightly above 100. These values may reflect service catchment areas that cross municipal boundaries or discrepancies between serviced dwellings in the numerator and total dwellings in the denominator.

levels. Specifically, we estimate an increase of 0.134 standard deviations in our index of public service provision (p-value < 0.01). The effects are positive for all components of the index and statistically significant in all but one case, ranging from 0.07 to 0.19 standard deviations across services. This suggests that mayors did not trade off performance in one task for another.

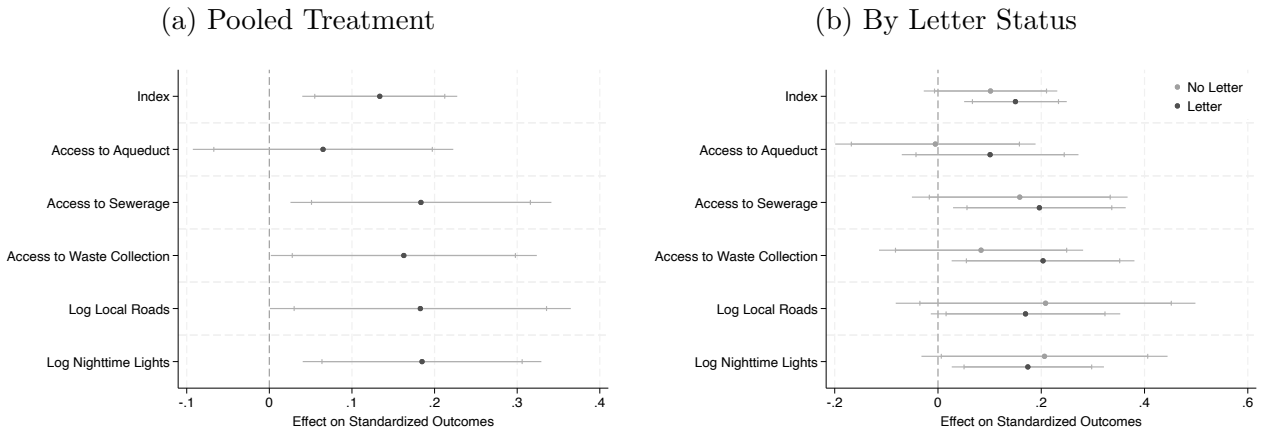
Panel (B) of Figure 1 reports estimates separately for municipalities where candidates received letters informing them about the intervention and those where they did not. We find larger and more precisely estimated effects in municipalities where candidates received the letter (0.15 SD, compared to 0.10 SD). This is also true for most components of the index, except for local roads and public lighting. This pattern mirrors the findings in Garbiras-Díaz and Montenegro (2022), where the letter treatment generated larger reductions in electoral irregularities, and is consistent with a link between electoral integrity and subsequent improvements in public service provision. However, the difference between these treatment arms is not statistically significant (p-value = 0.44 for the main index), so these results should only be taken as suggestive.

To assess the magnitudes of these effects, Appendix Table A4 reports estimates in the original units of each outcome. The intervention increased access to sewerage and waste collection services by 2.9 and 2.5 percentage points, respectively, while increasing our measures of local roads and public lighting by 13% and 3%. These are substantial improvements, representing increases of 5-7% in access to sewerage and waste collection relative to the control group's average at baseline, as well as increases of 1% and 10% in roads and public lighting. When expressed as differences between the baseline and the endline periods, the comparison is even starker, as improved access to sewerage and waste collection more than doubled the rate of increase in the period compared to the control group, while the rate of growth of local roads increased by 20%, and the rate of growth of public lighting increased by 10%.

These magnitudes are broadly comparable to those reported for other governance interventions affecting public services documented. In the water and sanitation sector, Galiani et al. (2005) find that the municipal-level privatization of water supply in Argentina increased access by 4.2 percentage points from a baseline of 64–86%. Similarly, using cross-country variation and an instrumental variable strategy, Cesar (2019) finds that a one-standard-deviation increase in the private ownership of water and sanitation services is associated with a 3–4.8 percentage point increase in water access.[31] Wong et al. (2017) show that self-governance reforms in Chinese villages increased the length of newly built roads by 44–

[31] These estimates come from scaling the 2SLS estimates in Table 4 (0.94-1.62) by the standard deviations of the outcome and instrumented variables reported in Table 1.

Figure 1: Effects on Public Service Index and Its Components



Note: This figure presents the effect of the intervention on the public service index and its components (expressed in z-scores, as described in Section 3.1). Panel (A) displays the effect of the pooled treatment arms, while Panel (B) displays the separate effects of sending the letter to candidates or not. These results correspond to Table A3. 95% and 90% confidence intervals are reported.

49%.[32] Likewise, Burgess et al. (2015) show that Kenyan districts sharing the president's ethnicity received five times more road construction than other districts, although this effect was present only during non-democratic periods. Finally, in the context of electrification, Baskaran et al. (2015) find that electoral cycles induced by special, unanticipated elections in India increase electricity provision by 4%, as measured by nighttime lights.

Dynamics. We next explore the dynamics of the effects in Figure A3, which reports the effects on the public service index for each year between 2020 and 2023. The estimated effect increases throughout the period, from just under 0.1 standard deviations in 2020 to approximately 0.15 standard deviations in 2023. This pattern is consistent with the idea that improvements in public service capacity take time to materialize. Figure A4 reports analogous estimates for each component of the index. We find a similar pattern across most outcomes, with somewhat steeper increases in waste collection and flatter trajectories for local roads.[33]

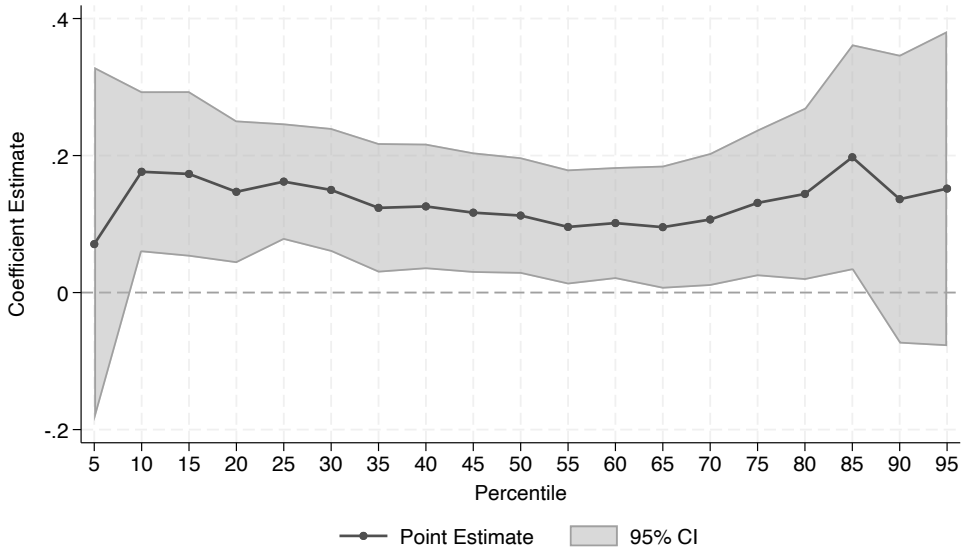
Distributional effects. While the previous results indicate a positive *average* treatment effect of the intervention, they may mask heterogeneity across the distribution. Figure A5 provides

[32] This estimate comes from expressing the estimates of self-governance reforms from Table 7 as a proportion of the mean of the outcome variable.

[33] The relatively flat trajectory for local roads is likely due to the crowdsourced nature of the data, which introduces several difficulties in pinpointing the exact date of construction of actual roads. Roads may be reported while still under construction, undergoing upgrades, or even at the planning stage, potentially making treatment effects to appear earlier than the completion of the underlying infrastructure projects.

a visual inspection of the cumulative distribution function (CDF) of the public service index by treatment condition, showing that the treatment shifted the distribution upward across most of its support. Consistent with this pattern, Figure 2 reports quantile treatment effects estimated at 5-percentage-point intervals. The effects are relatively stable throughout the distribution, with a modest increase around the 75th percentile and a decline at the 5th percentile.

Figure 2: Quantile Treatment Effects on the Public Service Index



Note: This figure presents the quantile treatment effects of the pooled treatment arms on the public service index by steps of 5 percentage points. Only strata fixed effects are used as controls. 95% confidence intervals are displayed.

Intra-municipal variation. We next examine heterogeneity between urban and rural areas. Theoretically, the ranking of the effects across these areas remains ambiguous. On one hand, economies of scale may lower the cost of infrastructure expansion in urban centers, incentivizing mayors to prioritize these areas, for instance, to maximize political returns from higher voter densities. On the other hand, rural areas typically start from lower levels of service provision (see Section 2), implying potentially larger gains from additional investments—a factor that would appeal to both altruistic and office-seeking mayors.

To investigate this question, we construct separate public service indices for urban and rural areas of each municipality.[34] The results, presented in Table A5, indicate that effects

[34] Administrative data on water, sewerage, and waste collection services allow for the disaggregation of properties by location type. For nighttime lights and road infrastructure, we use geographic coordinates to classify observations accordingly.

are slightly larger in urban areas than in rural areas (0.14 SD vs. 0.09 SD). Nevertheless, both estimates are statistically significant and economically meaningful.

Robustness. We conduct six different exercises to assess the robustness of our results. First, we examine the sensitivity of our findings to alternative data sources. As discussed in Section 3, our primary measures of aqueduct, sewerage, and waste collection rely on administrative records. While these cover the universe of municipalities in our experimental sample, they are subject to data quality limitations. As a robustness check, we turn to the household survey data described in Section 3. These data have the relative advantage of coming directly from public service users rather than being reported by providers. The results, presented in Appendix Table A6, confirm our main findings. Respondents in treated municipalities report higher access to public services, with an estimated effect of 0.08 standard deviations (p-value < 0.01) on an index aggregating the available outcomes. This magnitude is slightly smaller than, but similar to, the effect estimated using our main index. As before, effects are positive across all services and larger in the sub-treatment arm where letters were sent to candidates.[35]

Second, we examine whether our estimated effects on nighttime lights capture changes in underlying variables other than public lighting. One particular concern is that this variable might capture changes in economic activity, as documented in the development literature (e.g., Gibson et al. (2020)). To assess this possibility, we first examine whether the intervention affected municipal GDP over the studied period. As reported in column (1) of Table A7, we find no evidence that it did. We then residualize nighttime lights with respect to GDP and show that the intervention impacts only this residual component (columns (2) and (3) in Table A7). These results indicate that economic activity does not appear to be driving the results on nighttime lights.

Third, we assess the sensitivity of our findings to specific components of the public service index by re-estimating treatment effects while leaving out one component at a time. Figure A6 shows that the estimated effects remain positive, statistically significant, and similar in magnitude regardless of the omitted component.

Fourth, we examine whether the results depend on measuring outcomes as changes relative to the baseline period. To do so, we reconstruct the public service index using levels measured in 2022–2023 rather than changes between the pre- and post-intervention periods. As reported in Table A8, the estimated effects on the index and its components remain positive and mostly statistically significant. However, the magnitudes of the estimates decrease

[35] Interestingly, the effect on access to the aqueduct is now statistically significant at the 5% level, while access to the sewerage is not. This could be due to differences in the sample composition and the fact that survey respondents are not representative at the municipal level.

moderately, suggesting that differencing helps absorb residual baseline differences across municipalities.

Fifth, Table A9 explores the sensitivity of the results to including indicators for all treatment variants (e.g., by type of ad and type of letter sent to mayoral candidates), as recommended by Muralidharan et al. (2025). Remarkably, the estimates across all treatment arms suggest positive effects on the public service index. While the exact magnitude and significance vary across treatment arms, we cannot reject that they are all the same (p-value =0.21). This finding is consistent with an analogous exercise in Garbiras-Díaz and Montenegro (2022), which found limited evidence of meaningful differences across treatment variants, and supports our decision to pool treatment arms in the main analysis to increase statistical power.

Finally, Table A10 examines the sensitivity of the results to the inclusion of covariates as controls. As expected, the point estimates remain very similar, while standard errors increase modestly. The substantive conclusions are unchanged.

5 Mechanisms

In this section, we study the mechanisms through which enhanced electoral integrity improved access to local services. We first examine whether cleaner elections operated through the two channels emphasized in the political economy literature: the *selection* of politicians and their *accountability* once in office. To do so, we develop a simple framework that decomposes the overall effect into these channels and then provide evidence consistent with both. Next, we study how these channels might be reflected in concrete mayoral actions, focusing on procurement procedures and corruption-related investigations. Finally, we rule out alternative explanations related to fiscal capacity and support from higher levels of government.

5.1 Selection and Accountability Channels

Following seminal work in political economy on the welfare-enhancing features of democracy (Fearon 1999; Ferejohn 1986; Besley 2006), we posit that, by making elections cleaner, our intervention could have improved service provision through two channels: the election of politicians who are less involved in electoral irregularities (a selection channel) and increased effort by politicians once in office (an accountability channel).

To quantify the contribution of each channel to the overall effect estimated in the previous section, we develop a decomposition framework. We assume that the level of public service provision in municipality m when politician i is in office (Y_{mi}) is a function of three elements:

(i) the type of mayor in power $\phi_i \in \{clean, dirty\}$, (ii) this mayor's effort in providing these goods, e_i, and (iii) a municipality-specific term ξ_m, capturing all other determinants of public service provision that do not directly depend on the local politician in office, such as national programs, intergovernmental transfers, or citizen-led initiatives. We assume that this function is linear, so that it simplifies to the following expression:

$$Y_{mi} = \phi_i + e_i + \xi_m \tag{2}$$

We note a few important conceptual clarifications. The labels *clean/dirty* are meant to capture an underlying, time-invariant characteristic of the politician: their propensity or comparative advantage to pursue office by manipulating elections through irregularities. A *dirty* mayor is one disposed to such practices whereas a *clean* mayor is one who is not. We treat this disposition as a fixed trait, distinct from the effort e_i a mayor exerts once in office. Although this trait enters public service provision directly through ϕ_i, we do not impose a sign on this term, reflecting our agnosticism about whether clean or dirty politicians provide better public services. In practice, we acknowledge that a politician's "cleanliness" is likely correlated with other characteristics, including their socio-economic or political background, but also their psychological traits. For now, we take this type as given and return to its relationship with other characteristics below.

We denote the average treatment effect of the intervention as $\Delta Y_{mi} \equiv E[Y_{mi} \mid D = 1] - E[Y_{mi} \mid D = 0]$, which can then be expressed as the sum of three terms:

$$\underbrace{\Delta Y_{mi}}_{\text{ATE}} = \underbrace{\Delta \phi_i}_{\text{Selection Channel}} + \underbrace{\Delta e_i}_{\text{Accountability Channel}} + \underbrace{\Delta \xi_m}_{\text{Alternative Channels}} \tag{3}$$

The first term captures selection, the second accountability, and the third all other channels operating at the municipal level, m, that do not vary with politician i. The selection channel operates on the premise that, by making elections cleaner, the intervention may have changed the type of politician who is elected. In turn, the characteristics of politicians elected under cleaner electoral conditions can influence the provision of local public goods. Formally, it can be expressed as the product of the change in the probability that a clean candidate is elected and the expected effect of electing such a candidate on public service provision:

$$\Delta \phi_i = \Delta Pr[i = \text{clean}] \times (E[Y \mid i = \text{clean}] - E[Y \mid i = \text{dirty}]) \tag{4}$$

According to this expression, we can obtain an estimate of the overall selection channel by

separately estimating the two terms in (4). We can obtain an estimate of the accountability channel under the stronger assumption that $\Delta\xi_m = 0$, by computing $\Delta e_i = \Delta Y_{mi} - \Delta\phi_i$. In Section 5.2.1, we provide evidence consistent with this assumption by examining the main alternative channels through which the intervention could have improved public service provision independently of local politicians and showing that these channels are unlikely to be driving the results in our setting. However, we acknowledge that this constitutes only suggestive evidence, and therefore, any estimate of the accountability channel should be interpreted with appropriate caution.

Another caveat to keep in mind is that, given the linear separability assumption of the type and effort of mayors, the proposed estimate of the selection channel will also capture changes in politicians' effort that are driven by their underlying type. In this sense, the estimate of the accountability channel will only reflect changes in effort that are not attributable to changes in type. This does not undermine the validity of the exercise, but rather clarifies an important aspect of the interpretation of the proposed decomposition.

5.1.1 Estimating the Selection Channel

We estimate the selection channel in four steps.

1. Obtaining estimates of politician types. Prior to the intervention in 2019, we conducted an online pre-treatment survey to collect citizens' perceptions of which candidates were likely to engage in electoral irregularities in the upcoming elections. These responses allow us to construct empirical proxies for the politician types defined in the framework above. Full details of the survey and methodology are provided in Garbiras-Díaz and Montenegro (2022); here, we summarize only the key elements of this measurement approach.

In this survey, we asked respondents whether each candidate was likely to engage in the main forms of electoral irregularities observed in this context (e.g., vote-buying, illicit advertisements, voter intimidation, etc.). We aggregated these answers at the candidate level by computing the percentage of respondents who believe a candidate would engage in any electoral irregularity. We denote this Z_{mc}, for candidate c in municipality m. In Garbiras-Díaz and Montenegro (2022), we show that this measure is positively correlated ($\rho \approx 0.4$) with an independent measure of candidates' past malfeasance compiled by the local NGO *Fundación Paz y Reconciliación* (PARES) in a subset of 48 municipalities in our sample.

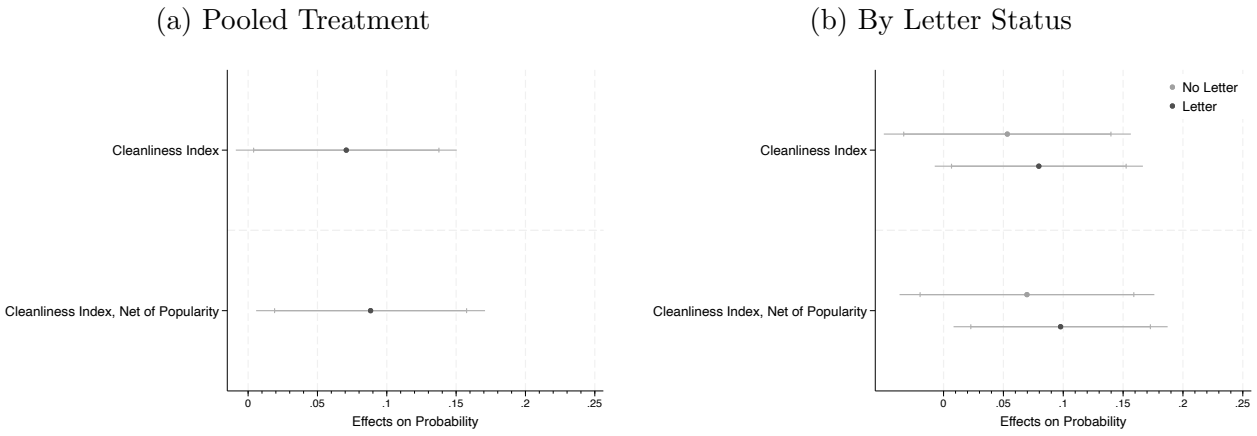
To maximize the use of available information, we construct an index of candidate cleanliness by combining these two measures. Namely, the first is the standardized version of Z_{mc}, reversed so that higher values correspond to cleaner candidates. The second is an indicator for whether candidate c was *not* included in PARES' list of candidates with past malfeasance.

We then define an indicator equal to one if this index is above the average across candidates municipality m. This indicator serves as our main measure of candidate cleanliness.

2. Estimating the effect of the intervention on electing clean mayors. Using this classification, we estimate $\Delta Pr\left[i = \text{clean}\right]$ by using the indicator for a "clean" mayor as the outcome in equation (1).

Figure 3 shows the main results (see column 1 of Appendix Table A11). While 31% of municipalities in the control group elected clean candidates, this propensity increased by 7.1 percentage points ($p < 0.1$) to 38% in the municipalities exposed to the intervention. This represents a large increase of 23% over the control group mean. Moreover, this effect is larger and more precise in municipalities in which politicians received a letter informing them of the online campaign, which is also what we find in Garbiras-Díaz and Montenegro (2022) with respect to the effects on electoral irregularities.

Figure 3: Effects on Probability of Electing a Clean Candidate as Mayor



Note: This figure presents the effects of the intervention on the probability that a clean candidate was elected. The outcome in the first row is an indicator for whether a clean candidate was elected according to the main index described in the text, while the outcome in the second row is an alternative measure that residualizes the popularity of candidates from this measure. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014); standard errors are clustered at the municipality level. The corresponding regression estimates in tabular form are reported in the first two columns of Table A11 (Appendix). 95% and 90% confidence intervals are reported.

In columns (3) and (4) of Table A11, we furthermore see that this finding is robust to different measures of candidate cleanliness—namely, when using a continuous version of Z_{mc} or an indicator for whether Z_{mc} is below the average in the municipality.[36]

[36]Panel A of Table A12 reports summary statistics for these three variables. Across the municipalities included in the analysis, the cleaner candidate was elected in approximately 37% of municipalities under both the standardized and raw cleanliness measures. When popularity is residualized out, the cleaner candidate is elected in 43% of cases.

3. Estimating the effect of electing a clean mayor on service provision. To estimate the second term in equation (4), $E[Y \mid i = \text{clean}] - E[Y \mid i = \text{dirty}]$, we exploit close elections in which the type of the winning candidate varies quasi-randomly while the accountability and alternative channels in equation (3) remain fixed.[37]

In practice, we implement a regression discontinuity design (RDD), using as the running variable the vote-share difference between the most successful clean candidate and the most successful dirty candidate. When this running variable is positive, a clean candidate wins the election; otherwise, a dirty candidate wins. More specifically, we estimate regressions of the following form:

$$y_m = \alpha \mathbb{1}\{R_m \leq 0\} + f(R_m) + e_m \tag{5}$$

where R_m is the running variable as defined before, and $f(R_m)$ is a polynomial of R_m which varies above and below the cutoff of $R_m = 0$.[38] This regression is estimated for observations that fall within a given bandwidth h of the cutoff. In our main specifications, we use the symmetrical optimal bandwidth proposed by Calonico et al. (2019). We use a triangular kernel and heteroskedasticity-robust nearest neighbor standard errors including the bias-robust correction by Calonico et al. (2014), but also report conventional p-values without this correction.

The key identification assumption is that potential outcomes are continuous around the cutoff of the running variable. Under this assumption, coefficient α identifies the local causal effect of electing a clean candidate rather than a dirty candidate in close races between the two types.[39] This assumption implies that, in close races between clean and dirty candidates, the election of a clean candidate is as good as random. We provide two main pieces of evidence in support of this assumption.[40] First, we show that the distribution of the running

[37] While the RD estimate is, by design, a local effect to a particular set of municipalities (e.g. the ones with barely winning/losing clean candidates), and is thus not representative of the more general effect of a clean candidate, this local effect is arguably the one of interest for this setting. The reason is that we are interested in knowing the effect of the intervention on the selection channel, which will arguably operate only in municipalities with competitive elections, where clean candidates were marginally elected due to the treatment.

[38] Panel B of Table A12 reports descriptive statistics for this running variable across the 595 municipalities for which the Cleanliness Index is defined. The average margin is −8 percentage points, although the running variable spans a wide range (from approximately −51 to 44 percentage points), providing substantial variation around the threshold.

[39] As with other politician-characteristic regression discontinuity designs, this effect may capture not only candidate cleanliness itself, but also other candidate attributes bundled with cleanliness (Marshall 2024). We return to this concern later in this section.

[40] These tests provide ex-post support for the identification assumption. A further, ex-post, theoretical argument supporting the assumption is that, although dirty candidates might manipulate elections, they cannot perfectly predict the non-manipulated voting counts, which means that the probability that they win

variable is continuous around the cutoff, suggesting the absence of manipulation in the vote share of clean candidates. Appendix Figure A7 presents the density test following Cattaneo et al. (2018). As shown in the figure, we fail to reject the null hypothesis of continuity at the threshold (p-value ≈ 0.4).

Second, we find no evidence of systematic discontinuities in municipal covariates around the cutoff. To assess this, we estimate equation (5) with each of the following as the outcome: baseline (2018–2019) values of the public service index, along with socio-economic covariates (GDP per capita, total population and density, rurality rates, poverty rates, and homicide rates), political variables (number of candidates, turnout, margin of victory, and votes in the 2014 presidential election), and other covariates (treatment assignment, Facebook penetration, and responses to the pre-treatment survey). Figure A8 presents the estimates of standardized covariates and their corresponding confidence intervals. Reassuringly, neither the point estimates for the public service index nor those for the other covariates are statistically significant at conventional levels. Despite these signs of balance in municipal characteristics, we also report specifications with covariates to increase the precision of estimates.

These findings lend support to the validity of the identifying assumption underpinning the regression discontinuity design. We now turn to the estimation results for equation (5), using the public service index as an outcome.

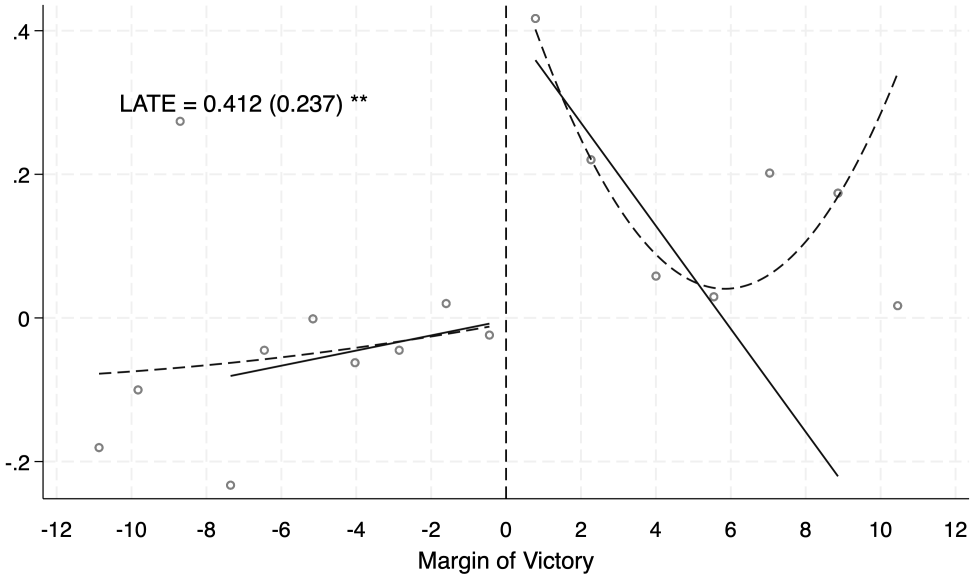
Figure 4 presents the main RDD results. Both linear and quadratic polynomials are fitted within their respective optimal bandwidths. The results consistently indicate that the election of a clean candidate leads to an increase in public service provision, regardless of the polynomial specification. The point estimate from the linear specification suggests that electing a clean candidate, as opposed to a dirty one, increases the public service index by a substantial 0.41 standard deviations (p-value < 0.05).

Columns (1)-(3) of Table 1 present the main point estimates, along with results from alternative specifications that vary the bandwidth by $\pm 30\%$ around the optimal value and employ both first- and second-order polynomials. Across all specifications, the estimated effects remain positive and statistically significant, reinforcing the robustness of the main findings. Notably, the main estimate of 0.41 standard deviations lies within the range of the alternative estimates, which vary between 0.36 and 0.59 standard deviations, depending on the specification.

Robustness. We conduct three robustness checks of the RD estimates. First, columns (4)-(6) of Table 1 include covariates to assess the sensitivity of the estimated effects. The estimates remain close to the original ones and statistically significant across different bandwidths and

is still as good as random when sufficiently close around the threshold, even though in the overall elections it is not.

Figure 4: RDD Estimating the Effect of Electing a Clean Candidate



Notes: This figure graphically illustrates the close-election regression discontinuity design on the public service index. It plots the binned average public service index according to the margin of victory of clean candidates, along with fitted linear and quadratic polynomials. The bandwidths used are the optimal ones of, approximately, 8.4 and 11.6 percentage points in the margin of victory. The estimated effect displayed corresponds to the one obtained with the linear polynomial. The standard error in parenthesis is the one estimated by the robust bias-corrected method proposed by Calonico et al. (2014).

polynomials used.

Second, we address the potential concern that the RD sample overlaps with the experimental sample, part of which was directly affected by the intervention. This could potentially introduce noise into the estimates, since public service provision was itself affected by the treatment, and it could also mask potential heterogeneous effects in the close-election RD estimates between treatment and control municipalities.

We address these concerns in two different ways. First, we control for treatment assignment in the baseline RD specification. As seen in column (1) of Table A13, this barely affects the point estimate and increases its precision. This is unsurprising given that treatment assignment is balanced around the cutoff (Figure A8). Second, we estimate the RD separately for treatment and control municipalities. The results, reported in columns (3) and (4) of Table A13, remain positive and similar in magnitude despite the loss of power associated with splitting the sample.

Third, we address the concern that survey responses about candidates' likely involvement in electoral irregularities may partly reflect candidate popularity rather than candidate clean-

liness.[41] To address this concern, we use a three-step approach. First, we use the share of survey respondents predicting that a candidate will win the election as a proxy for candidate popularity. Second, we residualize this measure from our cleanliness index using a linear regression. This measure would net out the systematic relationship between popularity and the cleanliness index in the survey.[42] Third, we re-estimate the RD using this residualized measure.

The results are shown in Table A14. The RD estimates remain positive, statistically significant, and virtually identical to the original ones, suggesting that popularity is not driving the results. Similarly, the second row of Figure 3 (and column (2) of Table A11) shows that the estimated effect of the intervention on the probability of electing a clean candidate is even larger and more precisely estimated using this alternative measure.

Table 1: RD Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. First Order Polynomial						
RD Estimate	0.412**	0.516	0.356**	0.401**	0.421	0.359**
	(0.204)	(0.375)	(0.173)	(0.159)	(0.557)	(0.156)
Bandwidth selection	Optimal	0.7 × Optimal	1.3 × Optimal	Optimal	0.7 × Optimal	1.3 × Optimal
Bandwidth	8.38	5.86	10.89	6.23	4.36	8.10
Controls				X	X	X
Conventional p-value	0.04	0.03	0.04	0.02	0.03	0.02
Bias-Robust p-value	0.04	0.17	0.04	0.01	0.45	0.02
Sample Size	209	161	265	169	120	207
Panel B. Second Order Polynomial						
RD Estimate	0.520**	0.594	0.440**	0.544***	0.459	0.476**
	(0.250)	(0.641)	(0.217)	(0.207)	(0.599)	(0.196)
Bandwidth selection	Optimal	0.7 × Optimal	1.3 × Optimal	Optimal	0.7 × Optimal	1.3 × Optimal
Bandwidth	11.57	8.10	15.05	10.20	7.14	13.27
Controls				X	X	X
Conventional p-value	0.04	0.05	0.05	0.01	0.05	0.01
Bias-Robust p-value	0.04	0.35	0.04	0.01	0.44	0.02
Sample Size	272	207	339	248	189	306

Notes: This table presents the point estimates of equation (5) using different bandwidths and polynomials on the change of the public service index. The specifications in columns (4)-(6) include the controls displayed in Figure A8. Standard errors in parenthesis are the ones estimated by the robust bias-corrected method proposed by Calonico et al. (2014); conventional and bias-robust p-values are shown in the lower part of the figure; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

4. Estimating the overall selection channel. Finally, combine the estimates obtained in Steps 2 and 3 according to equation (4) to estimate the contribution of the selection channel

[41] This relationship could arise in different ways. For instance, more popular candidates can be less accused of wrongdoing in virtue of being more popular among voters and respondents. Alternatively, more popular candidates could be easier to recognize by survey respondents making them more likely to be associated with any action, including irregularities.

[42] This procedure may also net out any genuine relationship between popularity and cleanliness that goes beyond the survey response bias described above. For instance, less popular candidates may be more likely to engage in electoral irregularities to remain competitive, in which case, it is unclear whether residualizing is desirable. For this reason, we continue to use the original cleanliness measure in our main analysis.

to the overall effect of the intervention. Bootstrapped p-values are displayed in Table A15.[43] The estimated contribution of the selection channel ranges from approximately 22% (p-value $\approx$ 0.07) in the baseline specification to 30% (p-value $\approx$ 0.04) when using the popularity-residualized measure of candidate cleanliness, suggesting that selection accounts for between one-fifth and one-third of the overall treatment effect.

Selection on other candidate characteristics. In the previous exercise, we focused on examining how the intervention impacted the selection of "clean" mayors. However, cleanliness is likely to be related to other candidate characteristics that may themselves be predictive of public service provision. In fact, as argued by Marshall (2024), regression discontinuity designs that study the effect of a particular politician characteristic risk bundling the attribute of interest with other correlated characteristics, thereby limiting the interpretability of the results.

To further shed light on the interpretation of our results, we now assess whether a candidate's cleanliness is correlated with other observable characteristics that might themselves affect outcomes. Ex-ante, the direction of these correlations is theoretically ambiguous. For instance, candidates more likely to engage in clientelism may also be more effective at delivering services because of their embeddedness in patronage networks (Kramon 2016), while candidates with criminal backgrounds may exhibit distinct patterns of behavior once in office (Khemka 2023).

We conduct three exercises. First, we examine how our measures of elected mayors' cleanliness correlate with a rich set of candidate characteristics, including demographic, educational, political, and professional attributes, as well as party affiliation, and sources of funding, drawing on the data sources described in Section 3. Second, we test whether these characteristics vary discontinuously at the RD threshold. Third, we examine whether the intervention itself affected the characteristics of elected mayors, beyond their cleanliness, in the 2019 elections.

Appendix Table A16 presents the correlation matrix between two alternative measures of cleanliness and the candidate characteristics described above. The correlations are small throughout ($\rho \leq 0.1$) and not statistically significant, with the exception of age, political experience, affiliation with a historical party, and campaign income.[44] Specifically, cleaner

[43] We compute p-values using a bootstrap procedure with $B = 1000$ replications. In each iteration, we re-estimate the ATE on the public service index, the effect of the intervention on the probability of electing a clean candidate, and the RD estimate. We then compute the implied selection effect using equation (4). The p-value is then computed using the bias-corrected empirical distribution of estimates, as suggested by Efron and Tibshirani (1993).

[44] Historical parties are traditional parties founded in the 19th century, when Colombia gained independence—namely, the Liberal and Conservative Parties.

candidates tend to be slightly younger, are less likely to have prior political experience, and report lower levels of campaign income.

Figure A9 then explores whether mayoral characteristics other than cleanliness exhibit systematic discontinuities at the threshold. Such differences would suggest alternative causal pathways through which electing clean candidates might affect public service provision. However, we find no systematic differences: of the 18 characteristics considered, only one exhibits a discontinuity statistically significant at the 10% level, which is consistent with what would be expected by chance alone.[45]

Finally, Tables A17-A19 examine whether the intervention affected the demographic, educational, political, professional, party, or campaign-finance characteristics of elected mayors. We find no evidence of systematic differences along any of these observable dimensions. In addition to the variables examined here, Garbiras-Díaz and Montenegro (2022) show that the intervention did not affect the probability of electing the most popular candidate, as measured in our pre-treatment citizen survey, a result that is also important for identifying the selection channel.

Taken together, these results indicate that the estimated selection effect is unlikely to be driven by the observable characteristics examined here. While these variables capture many of the candidate attributes most likely to shape public service provision, unobservable traits may still play a role. Attributes such as honesty or other unmeasured psychological determinants of performance could still be correlated with our measure of cleanliness and partially account for our findings. However, due to their inherently unobservable nature, assessing the contribution of these alternative factors remains outside of the scope of this paper.

5.1.2 Recovering the Accountability Channel

Having estimated the contribution of the selection channel, we now turn to the accountability channel. Under the assumption that the intervention did not affect public service provision through alternative channels unrelated to politicians, equation (3) implies that the accountability channel can be obtained as a residual: $[\Delta e_i = \Delta Y_{mi} - \Delta \phi_i]$.

As discussed below, we provide evidence consistent with this assumption by examining alternative channels related to fiscal capacity and support from higher levels of government (see Section 3.2). Our estimates therefore imply that the remaining 70–78% of the intervention's

[45] Notably, this attribute is one of the two measures of party incumbency we consider, while the estimate for the second measure is not statistically significant. The difference stems from their definitions: the latter strictly defines an incumbent party as identical to the one in power during the previous period. In contrast, the former adopts a broader definition, allowing for cases where the elected mayor's party forms a coalition with the previous ruling party, or was part of the previous mayor's governing coalition.

effect can be interpreted as attributable to accountability. However, this interpretation requires qualification due to a potential measurement error in the candidate-cleanliness proxy. Under classical measurement error, the estimates obtained in Steps 2 and 3 are likely to be attenuated, which would suggest that the estimated selection might be a lower bound of the actual effect, and the residual accountability component an upper bound.

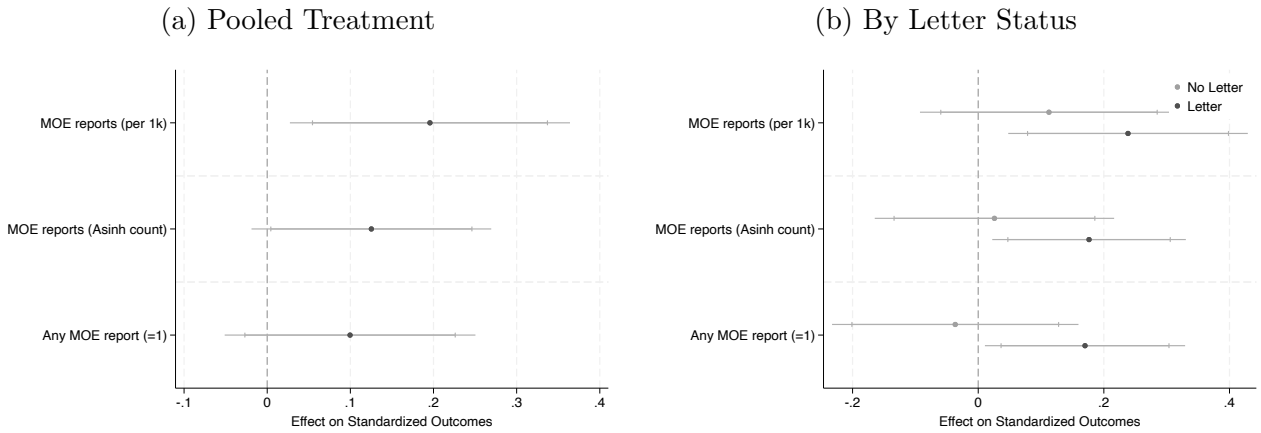
Electoral monitoring. How could the accountability channel have operated? A first possibility is through sustained electoral accountability that might have been spurred by the intervention. Although Colombian mayors cannot run for immediate reelection, future political careers, party incentives, and reputational concerns can discipline incumbents (see Section 2.1), and this will be more important if citizens continue to monitor future elections. We thus test whether the intervention produced lasting changes in citizens' electoral oversight, using reporting behavior during the 2023 local elections, which are the ones immediately following the 2019 ones.

Figure 5 reports the standardized treatment effects on three MOE-reporting outcomes: the number of reports per thousand citizens, the inverse hyperbolic sine transformation of the number of reports, and an indicator for whether any reports were filed in a given municipality.[46] Four years after the original intervention, treated municipalities continued to generate more reports. On the intensive margin, municipalities exposed to any treatments (Panel A) experienced a 0.2 SD increase in reports per thousand citizens (equivalent to 0.025 additional reports, $p < 0.05$), or a 0.13 SD increase in the asinh-transformed number of reports (equivalent to a 12% increase, $p < 0.1$). These estimates correspond to increases of 25% and 10.5%, respectively, relative to the control-group mean of each outcome. As expected, these effects are smaller than the one found in 2019, when the intervention increased reporting by 67% relative to the control-group mean, as reported in Garbiras-Díaz and Montenegro (2022). On the extensive margin, we find a positive but not statistically significant effect on the probability that a municipality files at least one report. This null result might be driven by the fact that, in 2023, 74% of municipalities filed at least one report, while this was only 29% in 2019. As before, we also find that the effects on all of these outcomes are larger, and statistically different in some cases, in municipalities where candidates were informed about the campaign than in municipalities where they were not (Panel B).

Non-electoral accountability. A second possibility is that the intervention increased *non-electoral* forms of accountability. Citizens can make demands to local governments outside the electoral arena either collectively or individually. We measure collective claims

[46] Table A20 reports the corresponding estimates in original units.

Figure 5: Effects on Reports to MOE in the 2023 Elections



Note: This figure presents the effect of the intervention on reports to MOE in 2023 local elections. The outcomes in each row are: the number of reports per thousand citizens, the inverse hyperbolic sine transformation of the number of reports, and an indicator for whether any reports were file in a given municipality, respectively. Panel (A) displays the effect of the pooled treatment arms, while Panel (B) displays the separate effects of sending the letter to candidates or not. Results for non-standardized outcomes are in Table A20. 95% and 90% confidence intervals are reported.

through protests and participation in grassroots organizations, as well as social media engagement, and individual judicial claims through constitutional writs (see Section 3.2). These actions may serve as vehicles for demanding more public services or greater government responsiveness.

We first test whether treated municipalities experienced more protests. The estimates, reported in Table A21, show no evidence that the intervention affected the incidence of protests per 10,000 inhabitants. Similarly, we find no effects focusing only on protests with a municipal scope, those demanding changes in public policies or those directed against the municipal executive government.

We also measure collective citizen demands through the presence and membership of *Juntas de Acción Comunal* (JACs), the most popular grassroots organizations in this context, which are often involved in supporting the provision of public services at the local level. In Table A22, we show the estimates on the number of JACs and on their membership per capita.[47] Across these different measures, we find small and not statistically significant effects.

Similarly, we examine whether the intervention more broadly shaped citizens' capacity to mobilize, proxied by social media engagement around political and social issues. We use the number of social media users and, in particular, the number of users who express interest

[47]For reference, we also report estimates for all grassroots organizations monitored by the Ministry of the Interior. Because the vast majority of these organizations are JACs, the estimates are nearly identical.

in political and social issues. We measure these variables as the log number of Meta users, as well as the number of users as a percentage of the population in each municipality. The estimates reported in Table A24 are again small and not statistically significant, suggesting that the intervention did not affect this proxy for citizen collective action.

Finally, we examine *tutelas*, constitutional writs that, as discussed in Section 3.2, are widely used by Colombian citizens to seek judicial protection of their constitutional rights, including access to basic public services. Table A23 reports the effects of the intervention on the number of writs filed per 1,000 inhabitants. Overall, we find no evidence that the intervention increased the number of constitutional writs filed, either in the aggregate or by type of claim. In particular, the estimated effect on writs demanding access to public works is small and not statistically significant.

All in all, we find no evidence that treated municipalities experienced greater citizen action outside the ballot box. These null results suggest that the accountability channel was more likely to operate through sustained electoral monitoring than through increased non-electoral mobilization.

5.1.3 On the Portability of the Decomposition

As a final remark about the decomposition of the selection and accountability channels, we note that it is not specific to our study of electoral integrity. We thus conclude by outlining the conditions under which it can be applied in other settings.

Three ingredients are essential. First, there must be an intervention whose total effect on the outcome is credibly identified, ideally through randomization. The second is a dichotomous candidate characteristic, observable prior to the election, that the intervention may affect by changing which types of candidates are elected. The third is a source of as-good-as-random variation in that characteristic among elected officials, used to estimate its causal effect on the outcome. In our setting, this variation comes from a regression discontinuity design in close races, although other designs could serve the same role.

Crucially, the variation used to estimate the effect of the characteristic must identify that effect for the same population whose selection is shifted by the intervention. The reason is that equation (4) combines two objects: the change in the probability of electing a given type induced by the intervention and the effect of that type on the outcome. If these two objects are estimated for different populations, the decomposition combines a selection effect from one population with the effect of the characteristic estimated in another, making the resulting estimate difficult to interpret. In our case, the two estimands coincide: the discontinuity identifies the effect of electing a clean candidate only in close races, and the intervention's selection channel likewise operates by flipping the outcome of close races.

Two additional assumptions are required for the decomposition to be informative. First, the intervention must move the selection margin: if it leaves the composition of elected types unchanged, there is no selection channel to estimate. Second, recovering the accountability channel as a residual requires both that the outcome be approximately additively separable in candidate type, effort, and factors outside the politician's control, as in equation (2), and that the intervention not affect the outcome through the municipal-level alternative channels in equation (3). These assumptions must be argued or bounded rather than presumed.

A final, more subtle concern, emphasized by Marshall (2024), is that close-election designs comparing candidates by a given characteristic may bundle that characteristic with other correlated attributes. A candidate who wins despite lacking the characteristic of interest may compensate along other dimensions that themselves affect the outcome. One then risks attributing the effect of those correlated attributes to the characteristic under study.

The severity of this concern depends on both the setting and the intervention, and in ours it is limited for two reasons. First, the characteristic entering the decomposition—candidate cleanliness—is the same dimension the intervention targets, namely electoral integrity. This matters because the object required for the decomposition is not the effect of cleanliness in the abstract, but the effect of the compositional change induced by the intervention. Were the studied characteristic and the intervention's target to differ, the concern would re-emerge: the decomposition could then misattribute to the studied characteristic an effect that in fact runs through the dimension the intervention actually moves.[48] Second, the bundle the discontinuity "prices" in our setting appears to be largely cleanliness itself: we find that cleanliness is largely uncorrelated with other observable characteristics, and that these neither vary discontinuously at the electoral threshold nor respond to the intervention—though we cannot rule out correlation with unobservable traits, including psychological ones. We therefore view this method as best suited to settings in which the studied characteristic coincides with the dimension the intervention targets, and we recommend that any application examine correlated candidate attributes directly, as we do here and as Marshall (2024) advises.

5.2 Mayors' Intermediate Actions

Having shown that both selection and accountability may have contributed to the overall effect, we now study actions by mayors that could have translated into improved public

[48] An example would be an intervention that encourages voters to support female candidates, paired with a decomposition based on whether candidates have been corrupt in the past. The dimension the intervention moves is gender; if past corruption is correlated with gender, the corruption-based decomposition would misattribute the effect of gender to corruption.

service delivery. We focus on two sets of outcomes that closely reflect mayoral decisions in the management of public resources and are particularly relevant for the provision of public services in our setting. The first, procurement practices, captures how local governments allocate public resources. The second, corruption-related investigations, captures whether mayors or their offices are associated with administrative misconduct.

Changes in these mayoral actions could be the consequence of either the election of cleaner mayors or greater effort by mayors once in office. As such, in addition to estimating the average effect of the intervention on these outcomes, we study how these effects relate to mayoral type and effort.

1. Procurement practices. In the Colombian context, public procurement has long been recognized as an area particularly prone to corruption and inefficiencies (Gulzar et al. 2022). It is also plays a key role in how Colombian mayors deliver services: procurement sits upstream in the service-delivery pipeline, and mayors have substantial discretion over these decisions.[49] We therefore test whether our intervention changed procurement practices in treated municipalities, focusing on the use of less competitive procurement modalities, which have been widely used in the literature as proxies for corruption risk and inefficiency (Fazekas et al. 2025; Baltrunaite et al. 2020).

Figure 6 and Table A25 summarize the treatment effects on these outcomes, described in Section 3.2. Estimates suggest that treated municipalities significantly increased their use of competitive procurement procedures (by 0.11 SD)—such as public tenders and open auctions—and decreased their reliance on non-competitive procedures, particularly direct contracting (by 0.16 SD). In original units (Table A25), governments in treated municipalities spent over 2 percentage points more of their budget on competitive contracts and approximately 3 percentage points less on non-competitive, direct contracts. These findings are robust to whether the outcome variable is defined as a share of contract value or as the log absolute value.[50] Reassuringly, we again find that these effects are stronger for the group of municipalities where candidates received the letter informing them about the intervention (Panel B). These shifts suggest greater transparency and efficiency in the use of public funds, potentially reducing opportunities for clientelism and rent-seeking.

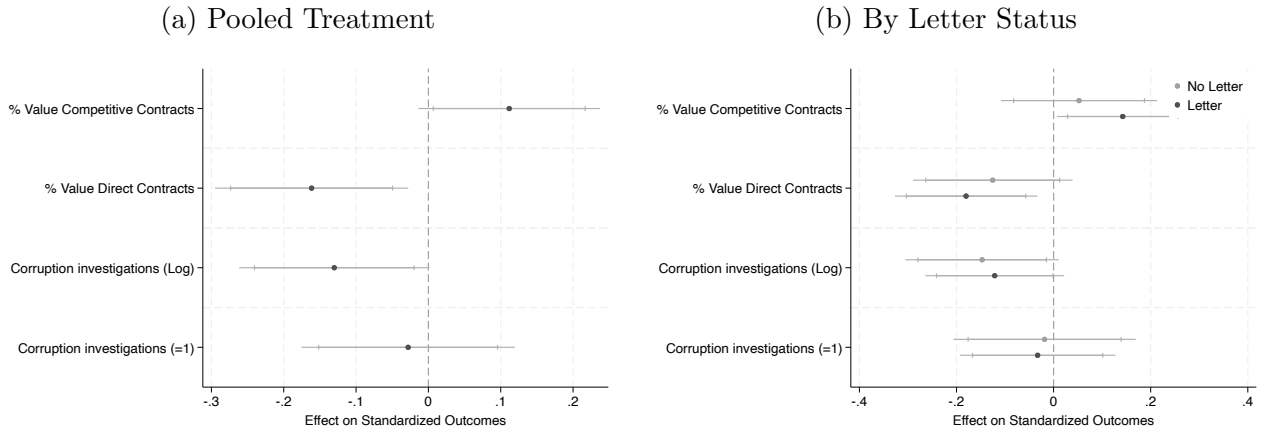
[49] Informal interviews with politicians and experts on local management confirm that procurement plays a central role in the provision of public services.

[50] We do not find that the total amount procured was affected by the intervention, as seen in column (1) of Table A25. This indicates that the effect is driven by a change in the modality of procurement, and not by differences in the amount of funds procured.

2. Corruption-related investigations. We then examine corruption in the public administration, which can reduce the resources that effectively reach citizens, distort the allocation of contracts, and lower the quality of public goods delivered (Colonnelli and Prem 2022; Olken 2007; Ferraz et al. 2012). Corruption investigations are especially relevant in this context because they are often triggered by irregularities in contract execution, public spending, or other administrative practices directly connected to local service delivery (Zuleta et al. 2018; OECD 2022). We thus measure the number of corruption-related judicial investigations against municipal governments over the study period.

Estimates are displayed in Figure 6 and Table A26. These results indicate that intervention led to a reduction in corruption-related judicial processes against mayoral offices in treated municipalities in the intensive margin (by 0.13 SD). Table A26 reports the corresponding magnitudes in original units: on the intensive margin, treated municipalities had 10% fewer corruption processes. Interestingly, we observe no such effect for judicial processes that do not have to do with corruption. On the extensive margin, treated municipalities also experience a decrease in the probability of having at least one corruption-related process, but this is not statistically significant. Since these legal proceedings are typically triggered by irregularities in contract execution and administrative practices, their decline provides further evidence of improved public management following the intervention.

Figure 6: Effects on Procurement and Corruption Outcomes



Note: This figure presents the effect of the intervention on procurement and corruption-related investigation outcomes, expressed as z-scores. The outcomes in the first and second rows are the percentage value of total contracts in 2020-2023 that are competitive and direct, respectively. The outcomes in the third and fourth row are the log plus one number of corruption investigations and an indicator for any such investigations in 2020-2023, respectively. Panel (A) displays the effect of the pooled treatment arms, while Panel (B) displays the separate effects of sending the letter to candidates or not. Results for non-standardized outcomes are in Tables A25 and A26. 95% and 90% confidence intervals are reported.

Effect of selection on mayoral actions. The previous analysis estimates the average effect of the intervention on mayors' procurement and corruption-related practices. As highlighted above, these effects could arise either because the intervention changed the type of mayor elected or because it increased the effort exerted by mayors once in office. To assess the extent to which selection drives these outcomes, we estimate the effect of electing a clean candidate using the close-election regression discontinuity design employed in Section 5.1.1.

Figure A10 and Table A27 report the estimates for the competitive-procurement outcomes and for corruption investigations. The results suggest that municipalities where a clean candidate barely won experienced fewer corruption-related investigations, but awarded a larger share of contract value through direct procurement and a smaller share through competitive procedures.[51] Quantitatively, barely electing a clean candidate reduces the number of corruption investigations by roughly 40% and the probability of facing one by 44 percentage points. At the same time, it increases the percentage of contract value awarded through direct procurement by 8.9 percentage points and lowers the percentage awarded through competitive procurement by 11.6 percentage points.[52]

Overall, these results suggest that the selection channel is an important driver of the reduction in corruption investigations induced by the intervention. By contrast, the evidence does not support selection as the driver of the observed increase in competitive procurement; if anything, it points in the opposite direction. This latter finding is a reminder that electing cleaner candidates need not be unambiguously positive for governance outcomes, even if, in net, they improve public services in their municipalities. Instead, it suggests that the increase in competitive procurement induced by the intervention is attributable to the accountability channel.[53]

5.2.1 Alternative Explanations

As a final exercise, we examine whether the observed gains in public service provision reflect additional inputs or actions by actors other than local governments, rather than a more efficient use of existing resources by mayors. We focus on two alternative mechanisms: increased fiscal capacity and additional support from higher levels of government.

[51] Figure A11 shows balance in the baseline values of these variables at the cutoff, suggesting that a potential imbalance is not driving the results.

[52] The reported estimates correspond to specifications that include controls. The estimates are similar in magnitude without controls, though standard errors are larger and the estimates lose significance for direct contracting and for the intensive-margin measure of corruption-related processes.

[53] Indeed, using equation (3), the accountability effect accounts for an increase of 3.08 percentage points in the percentage value of competitive contracts. This equals the overall effect of 2.26 percentage points minus the selection effect of -0.82 percentage points ($= 0.071 \times -11.6$). In other words, shutting down the selection channel would have implied a 36% larger increase in competitive procurement.

1. Increased fiscal capacity. The intervention may have increased the resources available to municipal governments through several channels. First, it could have improved mayors' ability to raise locally collected revenues, primarily through property taxes and fines. Second, it may have increased mayors' efforts to secure national-level transfers or support from higher levels of government.

We begin by examining administrative data on municipal revenues. The results, presented in Appendix Table A28, show no significant differences in total local government revenue between treated and control municipalities. Furthermore, this holds true when we separately examine locally collected tax revenue and intergovernmental transfers, which constitute the largest component of municipal revenue.

These findings suggest that the observed increase in access to public services was not driven by expanded public revenues. Instead, the results are more consistent with municipalities using resources more efficiently, in line with the earlier evidence on procurement practices and corruption.

2. Additional support from higher levels of government. We next examine whether the intervention affected services provided primarily by higher levels of government. As explained in Section 2.2, education and healthcare services in the municipalities studied are largely provided by departmental and national authorities. Examining these services therefore provides an indirect test of whether the observed improvements in public service provision were driven by differential support from higher levels of government. We test this possibility by examining differences in education and healthcare inputs, which lie directly outside the mayor's purview.[54]

We begin with education inputs, including the number of facilities, teachers, administrative staff, and students. The results, reported in Table A29, show no differences in any of these variables, whether measured in levels or relative to municipal population.

We then examine healthcare inputs, including the number of facilities, physicians, nurses and assistants, stretchers, beds, ambulances, offices and consultation rooms. The results, reported in Table A30, similarly show no overall effect of our intervention in these inputs (other than a small positive effect on the number of consultation rooms).

Altogether, this evidence suggests that the improvements in local public service provision generated by the intervention are more likely attributable to more efficient use of existing resources rather than to an increase in fiscal capacity or additional support from higher-order levels of government.

[54] We focus on inputs rather than outputs because the latter could reflect downstream effects of improvements in municipal services, such as sanitation or infrastructure.

6 Concluding Remarks

This paper provides novel experimental evidence on the causal relationship between electoral integrity and citizen welfare. By leveraging a randomized intervention that reduced electoral irregularities in Colombia's 2019 mayoral elections, we show that improvements in the quality of elections generate substantial benefits for public service provision. Furthermore, we find that these improvements were driven by better governance practices, as reflected in more competitive procurement processes and a reduction in corruption-related judicial processes.

Our findings carry significant policy implications. Most importantly, they highlight the importance of investing in mechanisms that guarantee free and fair elections as a means to promote local development. They bridge a long-standing "big picture" perspective that emphasizes the role of institutional quality in generating development endogenously (e.g., Acemoglu et al. 2005), with the view that relatively small but targeted interventions can unlock meaningful developmental gains (e.g., Duflo 2017). More concretely, the success of our light-touch intervention in producing substantial improvements suggests that citizen oversight mechanisms can be powerful tools for both strengthening democratic accountability and increasing downstream public good and service provision.

A natural question concerns the external validity of these results. The specific intervention we study—a low-cost social media campaign encouraging citizen monitoring—relies on conditions that may not hold everywhere (see Garbiras-Díaz and Montenegro (2022) for a detailed discussion of such conditions) or that may not be sustainable over time as political actors learn to manipulate elections in ways immune to citizen oversight. Nevertheless, the broad lessons gained from this rare instance of an intervention successfully increasing electoral integrity provide valuable insights into the tangible benefits of well-functioning democracies, as well as the costs societies incur when elections deviate from this ideal. These lessons resonate globally as democracies face growing challenges to electoral quality, not only from direct irregularities but also from eroding institutional legitimacy and the rise of populism (Guriev and Papaioannou 2022).

References

Acemoglu, Daron, Simon Johnson, and James A Robinson. 2005. "Institutions as a fundamental cause of long-run growth." *Handbook of economic growth* 1:385–472.

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson. 2019. "Democracy Does Cause Growth." *Journal of Political Economy* 127 (1): 47–100.

Aggarwal, Shilpa. 2018. "Do rural roads create pathways out of poverty? Evidence from India." *Journal of Development Economics* 133:375–395.

———. 2021. "The long road to health: Healthcare utilization impacts of a road pavement policy in rural India." *Journal of Development Economics* 151:102667.

Akhtari, Mitra, Diana Moreira, and Laura Trucco. 2022. "Political Turnover, Bureaucratic Turnover, and the Quality of Public Services." *American Economic Review* 112, no. 2 (February): 442–93.

Alt, James, Ethan Bueno de Mesquita, and Shanna Rose. 2011. "Disentangling Accountability and Competence in Elections: Evidence from U.S. Term Limits." *The Journal of Politics* 73 (1): 171–186.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–1495.

Anderson, Siwan, Patrick Francois, and Ashok Kotwal. 2015. "Clientelism in Indian Villages." *American Economic Review* 105, no. 6 (June): 1780–1816.

Aruoba, S. Borağan, Allan Drazen, and Razvan Vlaicu. 2019. "A Structural Model of Electoral Accountability." *International Economic Review* 60 (2): 517–545.

Asher, Sam, and Paul Novosad. 2020. "Rural Roads and Local Economic Development." *American Economic Review* 110, no. 3 (March): 797–823.

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

Auerbach, Adam Michael. 2020. *Demanding Development: The Politics of Public Goods Provision in India's Urban Slums.* Cambridge: Cambridge University Press.

Augsburg, Britta, Andrew Foster, Molly Lipscomb, Anna Tompsett, and Bansi Malde. 2025. "Sanitation Infrastructure." *VoxDevLit* 16 (1).

Baltrunaite, Audinga, Cristina Giorgiantonio, Sauro Mocetti, and Tommaso Orlando. 2020. "Discretion and Supplier Selection in Public Procurement." *The Journal of Law, Economics, and Organization* 37, no. 1 (May): 134–166.

Barrington-Leigh, Christopher, and Adam Millard-Ball. 2017. "The world's user-generated road map is more than 80% complete." *PLOS ONE* 12, no. 8 (August): 1–20.

Baskaran, Thushyanthan, Brian Min, and Yogesh Uppal. 2015. "Election cycles and electricity provision: Evidence from a quasi-experiment with Indian special elections." *Journal of Public Economics* 126:64–73.

Bazzi, Samuel, Masyhur Hilmy, Benjamin Marx, Mahvish Shaukat, and Andreas Stegmann. 2025. *It Takes a Village Election: Turnover and Performance in Local Bureaucracies.* Working Paper, Working Paper Series 33533. National Bureau of Economic Research.

Beath, Andrew, Fotini Christia, Georgy Egorov, and Ruben Enikolopov. 2016. "Electoral Rules and Political Selection: Theory and Evidence from a Field Experiment in Afghanistan." *The Review of Economic Studies* 83, no. 3 (April): 932–968.

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): 608–50.

Besley, Timothy. 2006. *Principled Agents? : The Political Economy of Good Government.* Oxford University Press.

Bishop, Sylvia, and Anke Hoeffler. 2016. "Free and fair elections: A new database." *Journal of Peace Research* 53 (4): 608–616. Accessed February 24, 2025.

Björkman, Martina, and Jakob Svensson. 2009. "Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda." *The Quarterly Journal of Economics* 124 (2): 735–769.

Blattman, Christopher, Horacio Larreguy, Benjamin Marx, and Otis Reid. 2026. "Eat Widely, Vote Wisely? Lessons from a Campaign Against Vote Buying in Uganda." Working paper.

Bosio, Erica, Simeon Djankov, Edward Glaeser, and Andrei Shleifer. 2022. "Public Procurement in Law and Practice." *American Economic Review* 112, no. 4 (April): 1091–1117.

Bowles, Jeremy, and Horacio Larreguy. 2025. "Who Debates, Who Wins? At-Scale Experimental Evidence on the Supply of Policy Information in a Liberian Election." *American Political Science Review* 119 (4): 1684–1703.

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel. 2015. "The Value of Democracy: Evidence from Road Building in Kenya." *American Economic Review* 105, no. 6 (June): 1817–51.

Calonico, Sebastian, Matias D Cattaneo, and Max H Farrell. 2019. "Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs." *The Econometrics Journal* 23, no. 2 (November): 192–210.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

Caputo-Leyva, Jorge Enrique. 2022. *¿Mandato del cielo? Dinastías políticas y desarrollo económico local en Colombia (Mandate from Heaven? Political Dynasties and Local Economic Development in Colombia).* Technical report 33. Available at SSRN: https://ssrn.com/abstract=4210553. Documento CEDE, August. https://doi.org/10.2139/ssrn.4210553.

Carreri, Maria. 2021. "Can good politicians compensate for bad institutions? Evidence from an original survey of Italian mayors." *The Journal of Politics* 83 (4): 1229–1245.

Casey, Katherine, Abou Bakarr Kamara, and Niccoló F. Meriggi. 2021. "An Experiment in Candidate Selection." *American Economic Review* 111, no. 5 (May): 1575–1612.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2018. "Manipulation Testing Based on Density Discontinuity." *The Stata Journal* 18 (1): 234–261.

Cesar, Sylvia. 2019. "PRIVATIZATION OF WATER: EVALUATING ITS PERFORMANCE IN THE DEVELOPING WORLD." *Annals of Public and Cooperative Economics* 90 (1): 5–23.

Chattopadhyay, Raghabendra, and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72 (5): 1409–1443.

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): 486–490.

Colagrossi, Marco, Domenico Rossignoli, and Mario A. Maggioni. 2020. "Does democracy cause growth? A meta-analysis (of 2000 regressions)." *European Journal of Political Economy* 61:101824. ISSN: 0176-2680.

Collier, Paul, and Pedro C. Vicente. 2012. "Violence, bribery, and fraud: the political economy of elections in Sub-Saharan Africa." *Public Choice* 153 (1/2): 117–147.

Colonnelli, Emanuele, and Mounu Prem. 2022. "Corruption and Firms." *The Review of Economic Studies* 89, no. 2 (March): 695–732.

Coviello, Decio, and Stefano Gagliarducci. 2017. "Tenure in Office and Public Procurement." *American Economic Journal: Economic Policy* 9, no. 3 (August): 59–105.

Dal Bó, Ernesto, and Frederico Finan. 2018. "Progress and Perspectives in the Study of Political Selection." *Annual Review of Economics* 10 (Volume 10, 2018): 541–575.

Dal Bó, Ernesto, Frederico Finan, Olle Folke, Torsten Persson, and Johanna Rickne. 2017. "Who Becomes A Politician?" *The Quarterly Journal of Economics* 132 (4): 1877–1914.

Doucouliagos, Hristos, and Mehmet Ali Ulubaşoğlu. 2008. "Democracy and Economic Growth: A Meta-Analysis." *American Journal of Political Science* 52 (1): 61–83.

Duflo, Esther. 2017. "The Economist as Plumber." *American Economic Review* 107, no. 5 (May): 1–26.

Dumas, Christelle, and Ximena Játiva. 2025. "Better Roads, Better Off? Evidence on Upgrading Roads in Tanzania." *The World Bank Economic Review* 39, no. 1 (February): 104–123.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D. Hyde, Craig McIntosh, Gareth Nellis, Claire L. Adida, Eric Arias, Clara Bicalho, Taylor C. Boas, Mark T. Buntaine, Simon Chauchard, Anirvan Chowdhury, Jessica Gottlieb, F. Daniel Hidalgo, Marcus Holmlund, Ryan Jablonski, Eric Kramon, Horacio Larreguy, Malte Lierl, John Marshall, Gwyneth McClendon, Marcus A. Melo, Daniel L. Nielson, Paula M. Pickering, Melina R. Platas, Pablo Querubín, Pia Raffler, and Neelanjan Sircar. 2019. "Voter information campaigns and political accountability: Cumulative findings from a preregistered meta-analysis of coordinated trials." *Science advances* 5 (7): eaaw2612.

Efron, Bradley, and Robert J. Tibshirani. 1993. *An introduction to the bootstrap.* Chapman & Hall.

Falleti, Tulia G. 2010. *Decentralization and subnational politics in Latin America.* Cambridge University Press.

Fazekas, Mihaly, Viktoriia Poltoratskaya, Marc Schiffbauer, and Bence Tóth. 2025. *Procuring Low Growth: The Impact of Political Favoritism on Public Procurement and Firm Performance in Bulgaria.* Technical report Working Paper No. 11085. World Bank.

Fearon, James D. 1999. "Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance." In *Democracy, Accountability, and Representation,* edited by

Adam Przeworski, Susan C. Stokes, and BernardEditors Manin, 55–97. Cambridge Studies in the Theory of Democracy. Cambridge University Press.

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." In *Economic Approaches to Political Behavior,* edited by Morris P. Fiorina and Kenneth A. Shepsle, 101–122. University of California Press.

Fergusson, Leopoldo, Carlos Molina, and Juan Felipe Riaño. 2018. "I sell my vote, and so what? Incidence, social bias, and correlates of clientelism in Colombia." *Economía* 19 (1): 181–218.

Ferraz, Claudio, and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *The Quarterly Journal of Economics* 123 (2): 703–745.

———. 2011. "Electoral Accountability and Corruption: Evidence from the Audits of Local Governments." *American Economic Review* 101, no. 4 (June): 1274–1311.

Ferraz, Claudio, Frederico Finan, and Diana B Moreira. 2012. "Corrupting learning: Evidence from missing federal education funds in Brazil." *Journal of Public Economics* 96 (9-10): 712–726.

Finan, Frederico, Benjamin Olken, and Rohini Pande. 2017. "Chapter 6 - The Personnel Economics of the Developing State." In *Handbook of Economic Field Experiments,* edited by Abhijit Vinayak Banerjee and Esther Duflo, 2:467–514. Handbook of Economic Field Experiments. North-Holland.

Fujiwara, Thomas. 2015. "Voting Technology, Political Responsiveness, and Infant Health: Evidence From Brazil." *Econometrica* 83 (2): 423–464.

Gagliarducci, Stefano, and Tommaso Nannicini. 2013. "Do Better Paid Politicians Perform Better? Disentangling Incentives from Selection." *Journal of the European Economic Association* 11 (2): 369–398.

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky. 2005. "Water for Life: The Impact of the Privatization of Water Services on Child Mortality." *Journal of Political Economy* 113 (1): 83–120. Accessed June 1, 2026.

Garbiras-Díaz, Natalia, and Mateo Montenegro. 2022. "All Eyes on Them: A Field Experiment on Citizen Oversight and Electoral Integrity." *American Economic Review* 112, no. 8 (August): 2631–68.

García-Sánchez, Miguel, and Sebastián Pantoja. 2015. "Incidencia del clientelismo según 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,* 291–313. MOE.

Gethin, Amory. 2023. *Revisiting Global Poverty Reduction: Public Goods and the World Distribution of Income, 1980-2022.* Working Paper, Working Paper Series N°2023/24. World Inequality Lab.

Gibson, John, Susan Olivia, and Geua Boe-Gibson. 2020. "NIGHT LIGHTS IN ECONOMICS: SOURCES AND USES." *Journal of Economic Surveys* 34 (5): 955–980.

Gottlieb, Jessica, and Katrina Kosec. 2019. "The Countervailing Effects of Competition on Public Goods Provision: When Bargaining Inefficiencies Lead to Bad Outcomes." *American Political Science Review* 113 (1): 88–107.

Guerra, Alice, and Mogens K. Justesen. 2022. "Vote buying and redistribution. Public Choice." *Public Choice* 193:315–344.

Gulzar, Saad, and Muhammad Yasir Khan. 2024. "Good Politicians: Experimental Evidence on Motivations for Political Candidacy and Government Performance." *The Review of Economic Studies* 92, no. 1 (March): 339–364.

Gulzar, Saad, Miguel R Rueda, and Nelson A Ruiz. 2022. "Do campaign contribution limits curb the influence of money in politics?" *American Journal of Political Science* 66 (4): 932–946.

Guriev, Sergei, and Elias Papaioannou. 2022. "The political economy of populism." *Journal of Economic literature* 60 (3): 753–832.

Habyarimana, James, Macartan Humphreys, Daniel N Posner, and Jeremy M Weinstein. 2009. *Coethnicity: Diversity and the dilemmas of collective action.* Russell Sage Foundation.

Hidalgo, F. Daniel, and Simeon Nichter. 2016. "Voter Buying: Shaping the Electorate through Clientelism." *American Journal of Political Science* 60 (2): 436–455.

International IDEA. 2024. *The Global State of Democracy 2024: Strengthening the Legitimacy of Elections in a Time of Radical Uncertainty.* Technical report. Stockholm: International Institute for Democracy and Electoral Assistance. https://www.idea.int/gsod/2024/.

Karim, Ridwan Mohammad. 2025. "Voter-buying, politician selection, and public good provision in Brazil." *Journal of Development Economics* 176:103507.

Khandker, Shahidur R., Zaid Bakht, and Gayatri B. Koolwal. 2009. "The Poverty Impact of Rural Roads: Evidence from Bangladesh." *Economic Development and Cultural Change* 57 (4): 685–722. Accessed August 25, 2025.

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

Khemka, Abhinav. 2023. "Manipulating the System: Clientelism and Criminality in Politics." *Available at SSRN 4995257.*

Kramon, Eric. 2016. "Electoral handouts as information: Explaining unmonitored vote buying." *World Politics* 68 (3): 454–498.

Londoño-Vélez, Juliana, and Estefanía Saravia. 2025. "The Impact of Being Denied a Wanted Abortion on Women and Their Children*." *The Quarterly Journal of Economics* (January).

Marshall, John. 2024. "Can close election regression discontinuity designs identify effects of winning politician characteristics?" *American Journal of Political Science* 68 (2): 494–510.

Martinez-Bravo, Monica, Gerard Padró i Miquel, Nancy Qian, and Yang Yao. 2022. "The Rise and Fall of Local Elections in China." *American Economic Review* 112, no. 9 (September): 2921–58.

Marx, Benjamin, Vincent Pons, and Vincent Rollet. 2024. "Electoral Turnovers." *The Review of Economic Studies* (November): rdae108. ISSN: 0034-6527.

MOE. 2023. *Mapa de Riesgo Electoral: Elecciones de Autoridades Locales 2019.* https://www.moe.org.co//wp-content/uploads/2019/11/Mapas-de-Riesgo-Electoral-2019-_DIGITAL.pdf. Accessed: June 6, 2025.

Muralidharan, Karthik, Paul Niehaus, Sandip Sukhtankar, and Jeffrey Weaver. 2021. "Improving Last-Mile Service Delivery Using Phone-Based Monitoring." *American Economic Journal: Applied Economics* 13 (2): 52–82.

Muralidharan, Karthik, Mauricio Romero, and Kaspar Wüthrich. 2025. "Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments." *The Review of Economics and Statistics* 107 (3): 589–604.

Norris, Pippa, Richard W. Frank, and Ferran Martínez i Coma. 2014. "Measuring Electoral Integrity around the World: A New Dataset." *PS: Political Science & Politics* 47 (4): 789–798.

OECD. 2022. *OECD Economic Surveys: Colombia 2022.* Technical report. Paris: OECD Publishing. https://doi.org/10.1787/04bf9377-en.

Ofosu, George Kwaku. 2019. "Do Fairer Elections Increase the Responsiveness of Politicians?" *American Political Science Review* 113 (4): 963–979.

Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (2): 200–249.

Paniagua, Victoria. 2022. "When clients vote for brokers: How elections improve public goods provision in urban slums." *World Development* 158:105919. ISSN: 0305-750X.

Przeworski, Adam. 2000. *Democracy and development: Political institutions and well-being in the world, 1950-1990.* 3. Cambridge University Press.

Sieg, Holger, and Chamna Yoon. 2017. "Estimating Dynamic Games of Electoral Competition to Evaluate Term Limits in US Gubernatorial Elections." *American Economic Review* 107 (7): 1824–57.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–586.

Superservicios. 2024. *Informe Sectorial de los Servicios Públicos Domiciliarios de Acueducto y Alcantarillado.* Technical report. Vigencia 2023. Superintendencia de Servicios Públicos Domiciliarios (Superservicios), República de Colombia, enero. https://www.superservicios.gov.co/sites/default/files/inline-files/Informe-Sectorial-de-los-Servicios-Publicos-Domiciliarios-de-Acueducto-y-Alcantarillado-Vigencia-2023.pdf.

Welsh, Brandon C., David P. Farrington, and Stephen Douglas. 2024. "The Impact and Policy Relevance of Street Lighting for Crime Prevention: A Systematic Review Based on a Half-Century of Evaluation Research." Https://www.crimrxiv.com/pub/wl9zqxga, *CrimRxiv* (April).

Wong, Ho Lun, Yu Wang, Renfu Luo, Linxiu Zhang, and Scott Rozelle. 2017. "Local governance and the quality of local infrastructure: Evidence from village road projects in rural China." *Journal of Public Economics* 152:119–132.

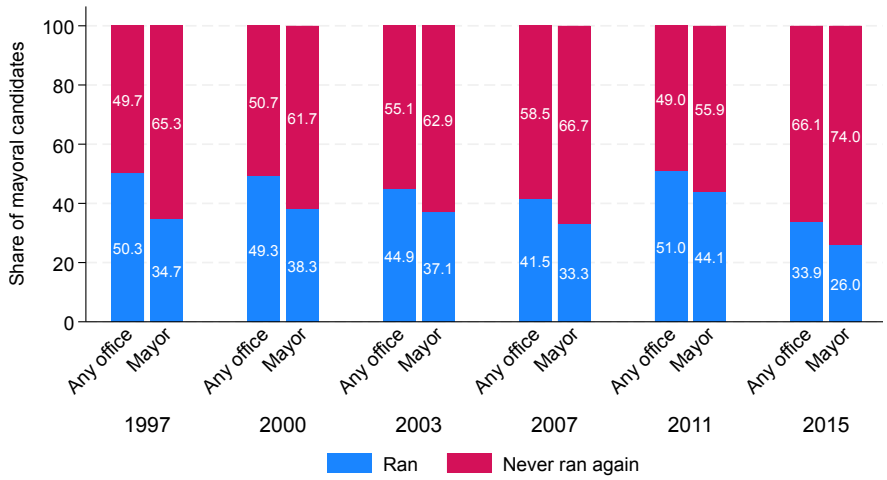
Young, Alwyn. 2018. "Channelling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." *he Quarterly Journal of Economics* 134 (2): 557–598.

Zuleta, María Margarita, Víctor Saavedra, and Juan Camilo Medellín. 2018. *Fortalecimiento del sistema de compra pública para reducir el riesgo de corrupción.* Technical report. Bogotá, Colombia: Fedesarrollo.

Supplementary Data and Appendix

A Appendix Figures

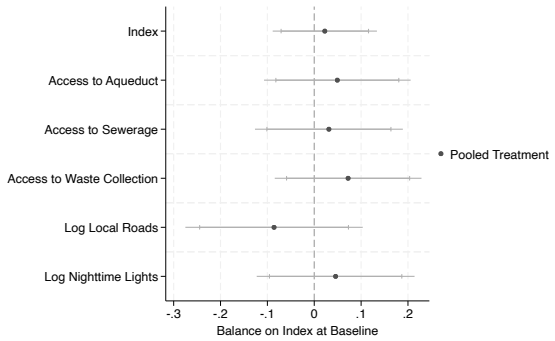
Figure A1: Elected mayors' subsequent candidacies by election cohort



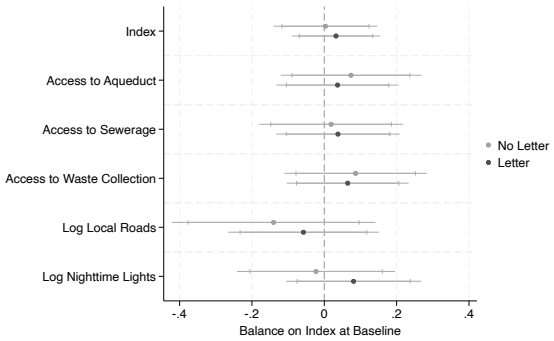
Note: The figure plots the share of mayors elected in each cohort who later ran for any subsequent elected office. Eligible follow-up contests are all local and national elections held between 2000 and 2023, excluding the presidency. Colombia holds local elections (mayor, municipal council, governor, departmental assembly) every four years in odd years and national elections (Senate, House) every four years in even years. *Any office* captures mayors who ran for any elected position; *Mayor* restricts to those who ran again for mayor in the same municipality in a later, non-consecutive term. Mayors face a one-term cool-off before rerunning for mayor and cannot run in the immediately following legislative elections unless they resign, so they can first reappear only in the next legislative cycle; the last observable cohort is therefore 2015. Source: *Registraduría Nacional del Estado Civil.*

Figure A2: Baseline Balance: Public Service Index and Mechanism Outcomes

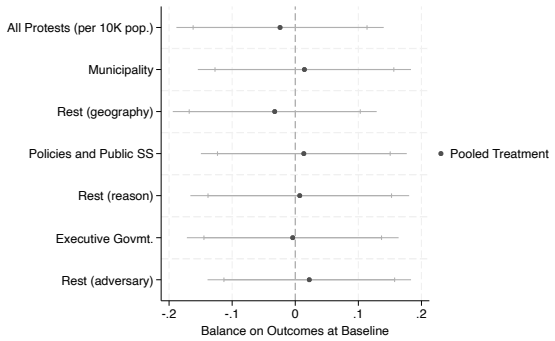
(a) Public Service Index & components



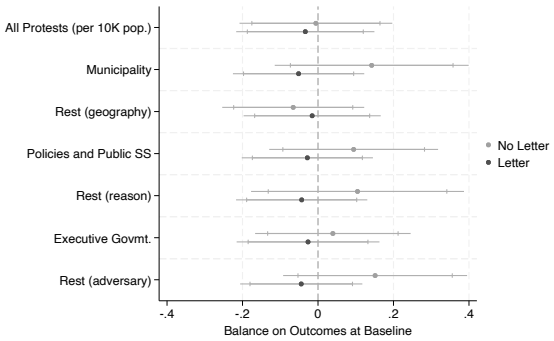
(b) Index & components — Letter vs. No Letter



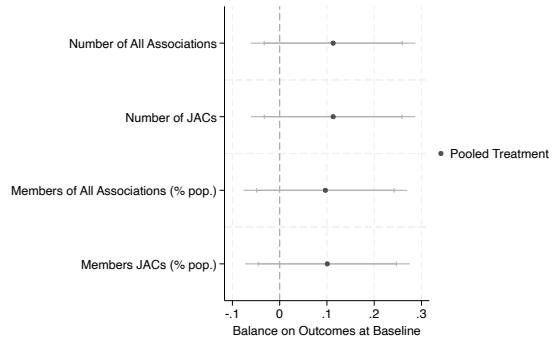
(c) Protests



(d) Protests — Letter vs. No Letter



(e) Grassroots associations



(f) Associations — Letter vs. No Letter

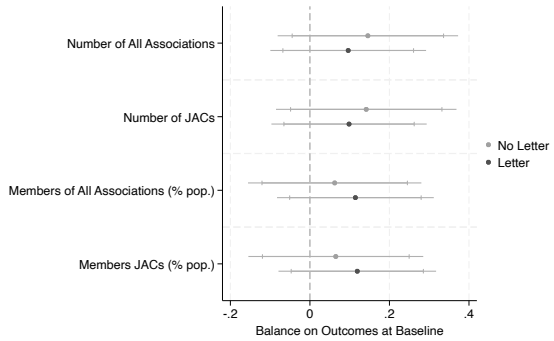
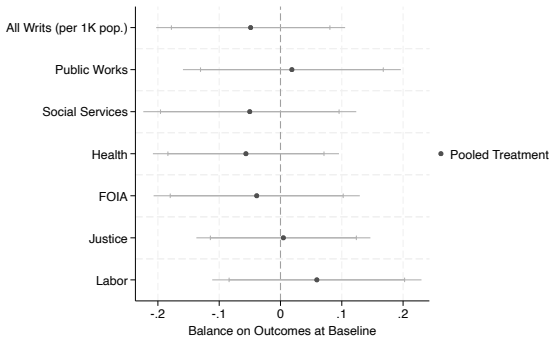
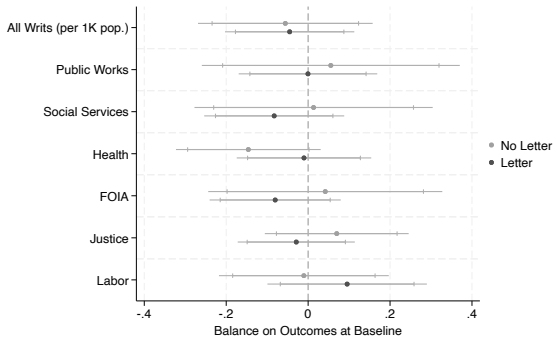


Figure A2: Baseline Balance (continued)

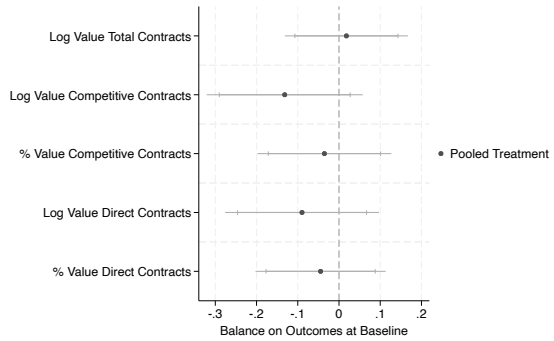
(g) Citizen writs (*tutelas*)



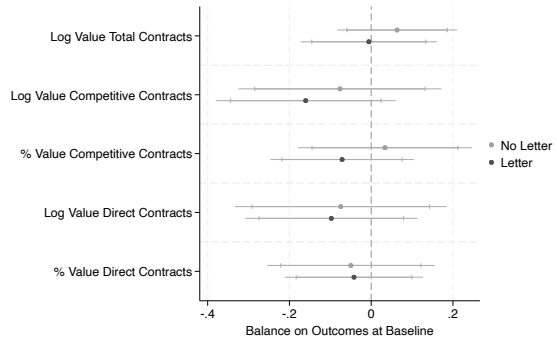
(h) Writs — Letter vs. No Letter



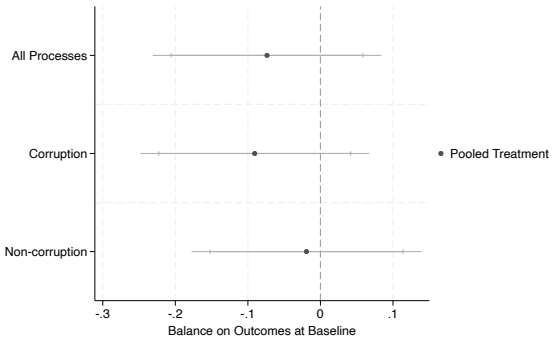
(i) Procurement



(j) Procurement — Letter vs. No Letter



(k) Judicial processes



(l) Processes — Letter vs. No Letter

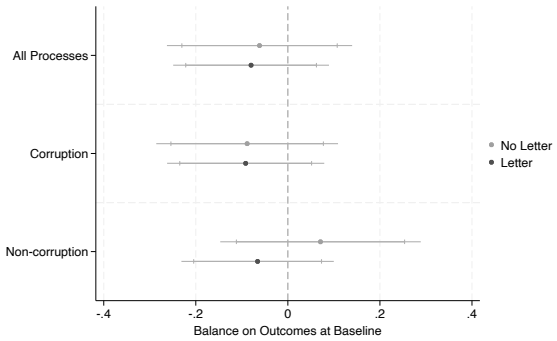
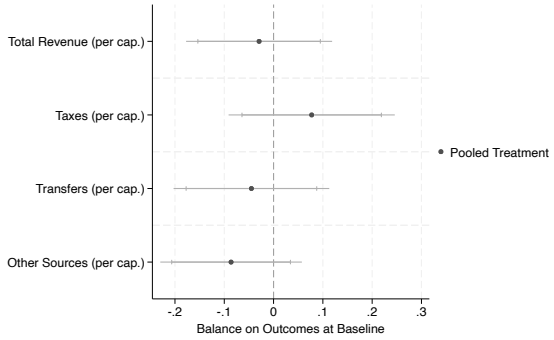
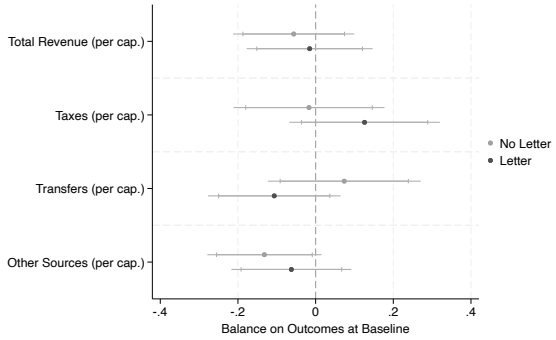


Figure A2: Baseline Balance (continued)

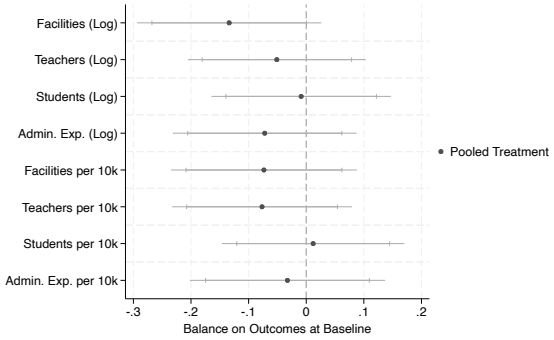
(m) Local public revenue



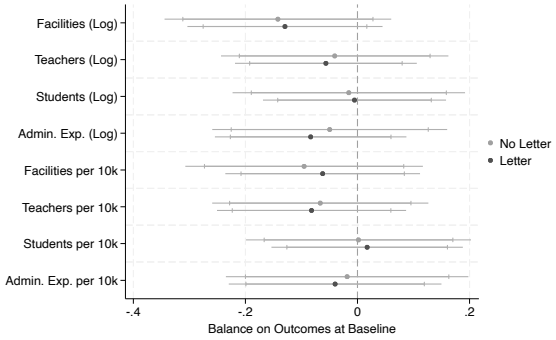
(n) Revenue — Letter vs. No Letter



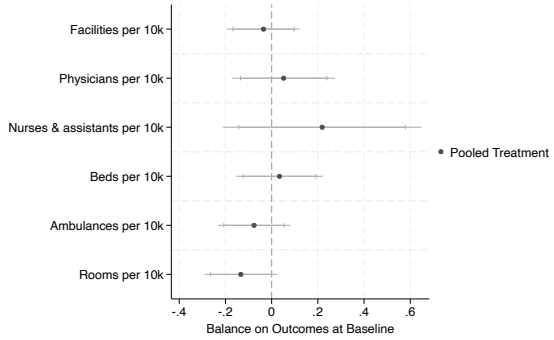
(o) Education



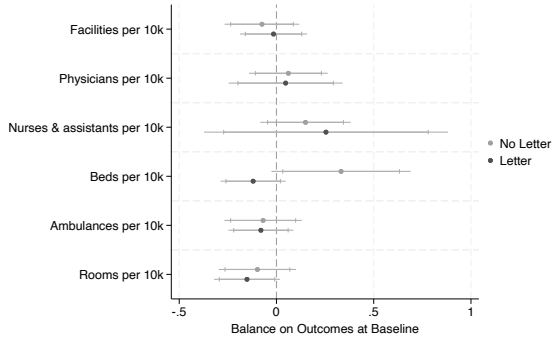
(p) Education — Letter vs. No Letter



(q) Healthcare provision

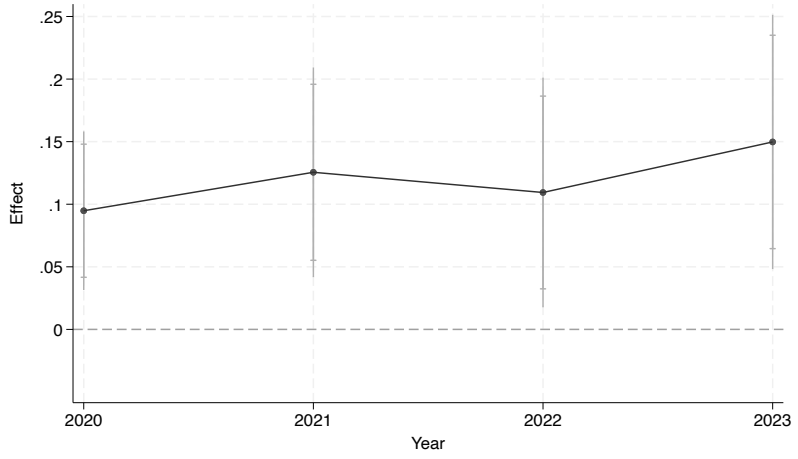


(r) Healthcare — Letter vs. No Letter



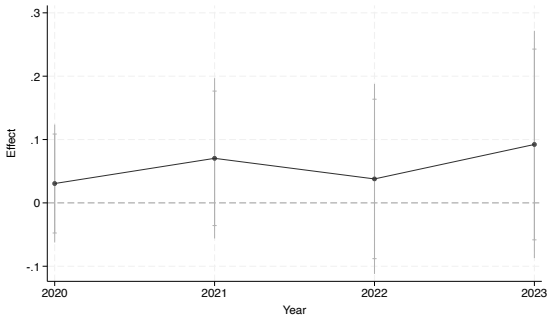
Note: This set of figures tests for the balance of the intervention on different baseline covariates (expressed as z-scores). Panel (A) displays the effect of the pooled treatment arms, while Panel (b) displays the separate effects of sending the letter to candidates or not. Two healthcare baselines (stretchers and consultation offices per 10k) are omitted because their 2019 pre-period value is missing, and Facebook-user outcomes are omitted because they are measured only post-treatment. 95% and 90% confidence intervals are reported.

Figure A3: Dynamic Effects on Public Service Index

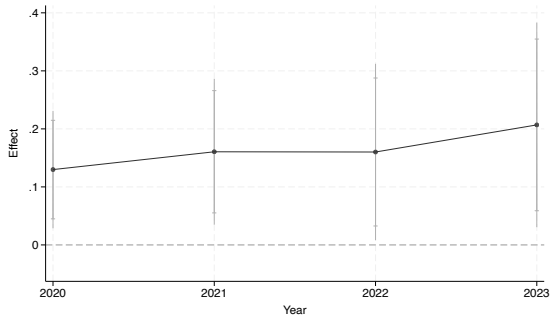


Notes: This figure displays the estimates of the intervention separately for each year in the study period on the public service index. 95% and 90% confidence intervals are reported.

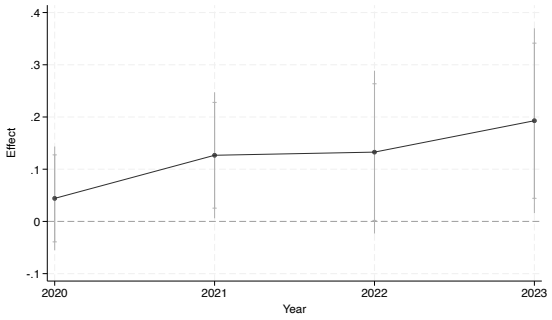
Figure A4: Dynamic Effects on the Components of the Public Service Index



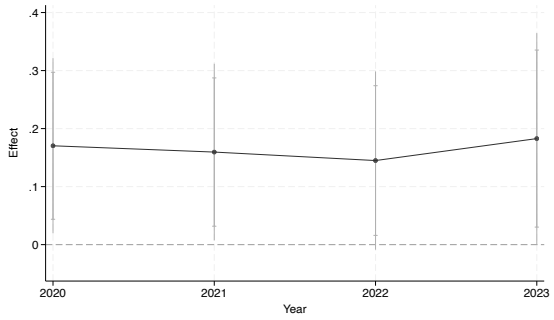
(a) Access to Aqueduct



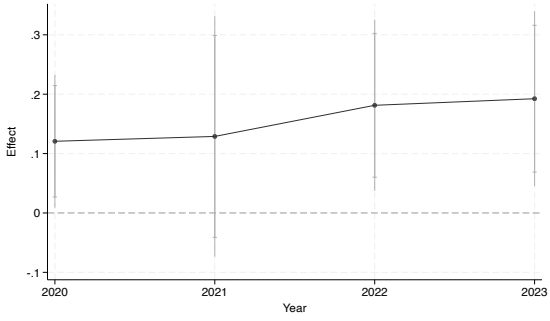
(b) Access to Sewerage



(c) Access to Waste Collection



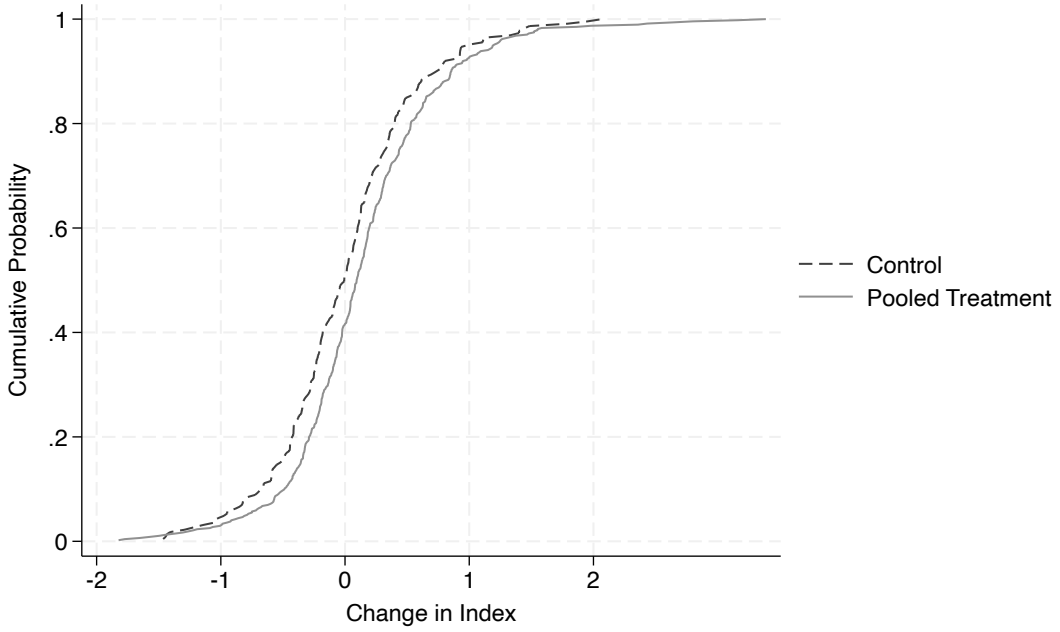
(d) Log Local Roads



(e) Log Nighttime Lights

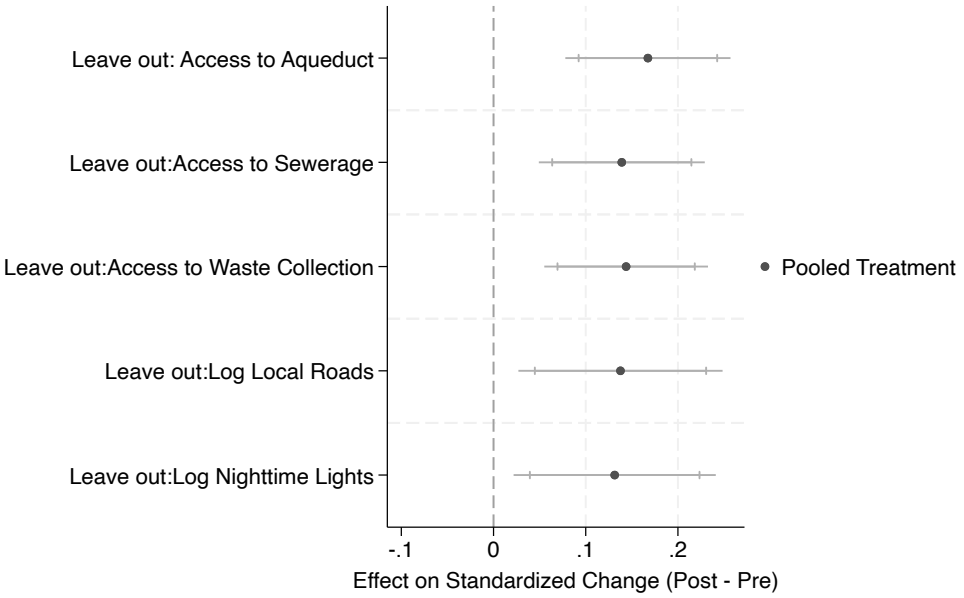
Notes: This figure displays the estimates of the intervention separately for each year in the study period on the different components of the public service index. 95% and 90% confidence intervals are reported.

Figure A5: Cumulative Distribution Function for Public Service Index by Treatment Condition



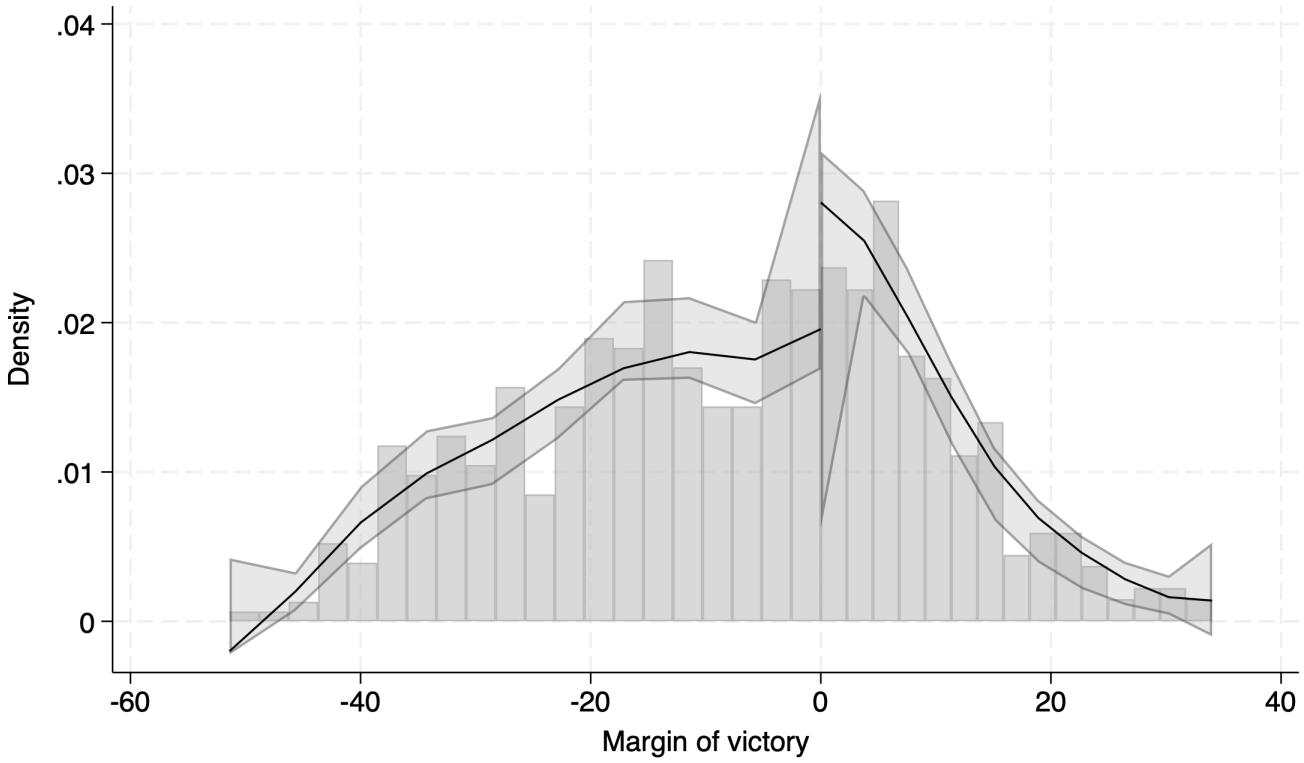
Note: This figure presents the cumulative density function of the Public Service Index, by treatment condition.

Figure A6: Leave-one-out Variations of Public Service Index



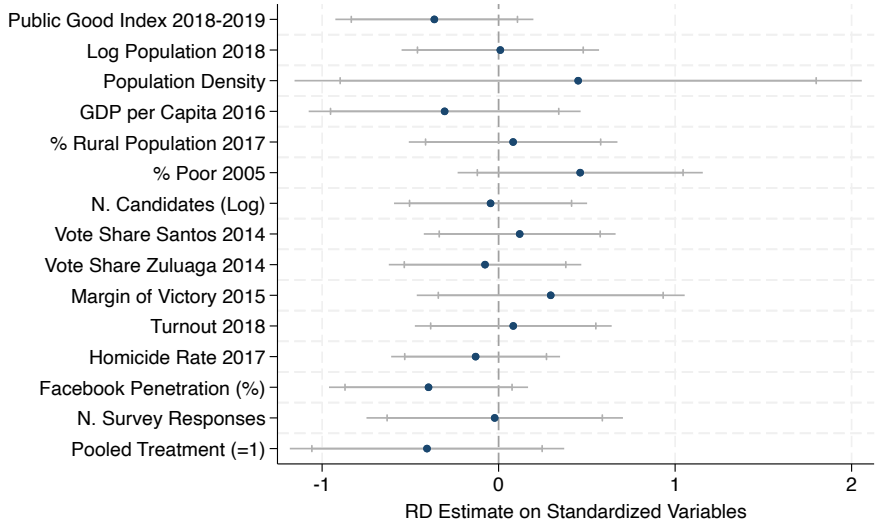
Note: This figure presents the estimates of the effects of the pooled treatment arms on variations of the public service index, when dropping one of the components at a time. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). 95% and 90% confidence intervals are reported.

Figure A7: RD Manipulation Test



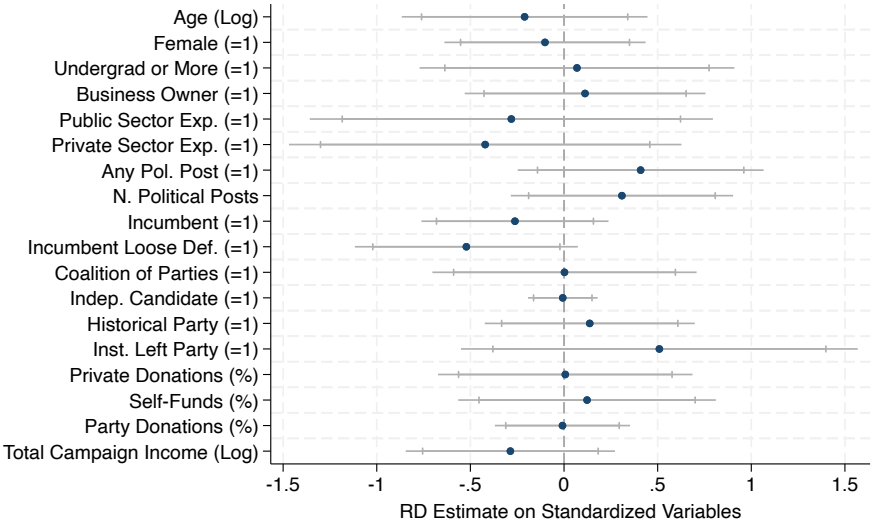
Notes: This figure displays the test for the continuity of distribution of the running variable in equation (5) proposed by Cattaneo et al. (2018).

Figure A8: RD Continuity of Covariates at Threshold



Notes: This figure displays the estimates of equation (5) using different covariates as outcome variables. All covariates were transformed into z-scores so the plotted effects can be interpreted as standard deviation differences. Standard errors are the ones estimated by the robust bias-corrected method proposed by Calonico et al. (2014). 95% and 90% confidence intervals are reported.

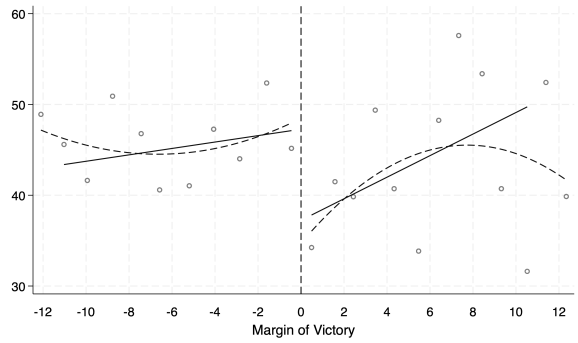
Figure A9: RD Candidate Characteristics at Threshold



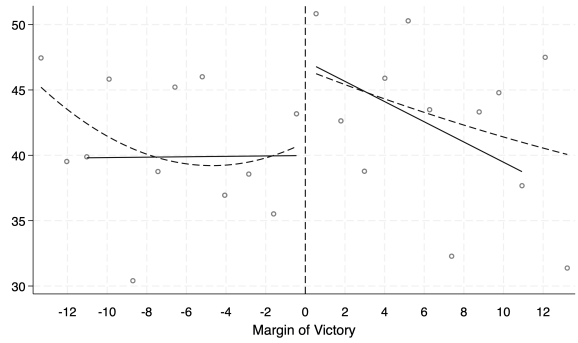
Notes: This figure displays the estimates of equation (5) using variables capturing different candidate characteristics as outcomes. All variables were transformed into z-scores so the plotted effects can be interpreted as standard deviation differences. Standard errors are the ones estimated by the robust bias-corrected method proposed by Calonico et al. (2014). 95% and 90% confidence intervals are reported.

Figure A10: RD Estimates on Mayoral Actions

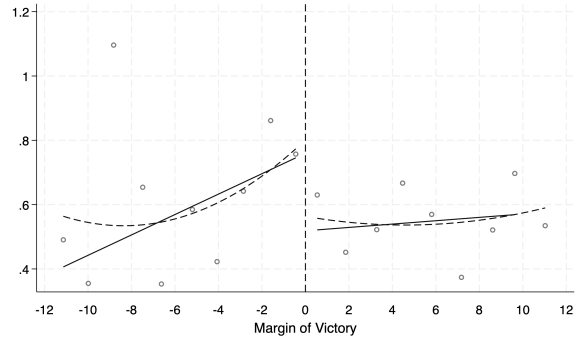
(a) % Value Competitive Contracts



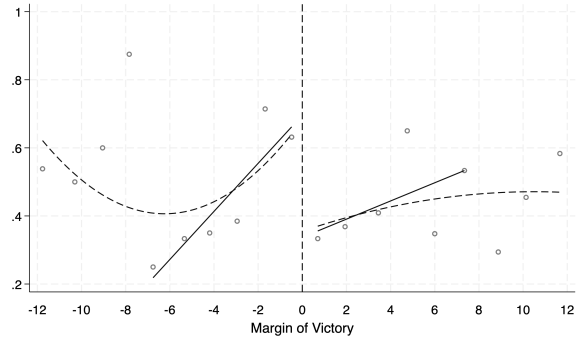
(b) % Value Direct Contracts



(c) Corruption Investigations (Log)

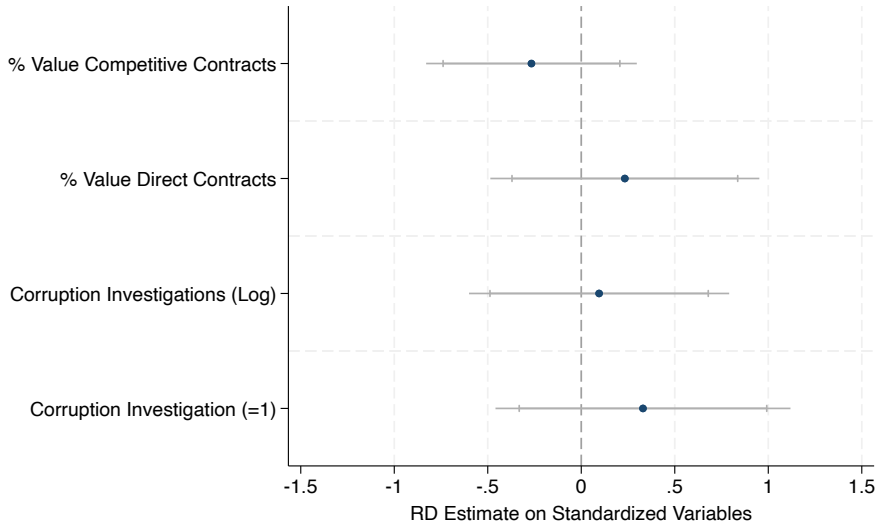


(d) Corruption Investigations (=1)



Notes: This set of figures graphically illustrate the close-election regression discontinuity design on the mayors' governance-related outcomes. It plots the binned average outcomes according to the margin of victory of clean candidates, along with fitted linear and quadratic polynomials. The bandwidths used in each figure are the optimal ones.

Figure A11: RD Continuity of Baseline Mayoral Actions at Threshold



Notes: This figure displays the estimates of equation (5) using the different measures for mayors' governance-related outcomes measured in 2016-2019. All outcomes were transformed into z-scores so the plotted effects can be interpreted as standard deviation differences. Standard errors are the ones estimated by the robust bias-corrected method proposed by Calonico et al. (2014). 95% and 90% confidence intervals are reported.

B Appendix Tables

Table A1: Descriptive Statistics: Main Outcomes (Index and Subcomponents)

	N	Mean	SD	P25	P50	P75	Min	Max
Panel A. Standardized Changes (post − pre)								
Index	698	0.10	0.63	-0.25	0.06	0.40	-1.82	3.39
Access to Aqueduct	695	0.04	0.98	-0.31	-0.09	0.46	-3.72	4.89
Access to Sewerage	695	0.12	0.98	-0.22	-0.01	0.49	-3.52	5.55
Access to Waste Collection	695	0.11	1.02	-0.21	0.03	0.51	-4.64	6.01
Local Roads (log)	698	0.09	1.26	-0.46	-0.09	0.36	-11.69	15.42
Nighttime Lights (log)	698	0.12	1.18	-0.49	0.11	0.74	-8.47	13.65
Panel B. Raw Changes (post − pre, levels)								
Access to Aqueduct	695	4.75	17.49	-1.69	2.32	12.12	-62.70	91.64
Access to Sewerage	695	3.93	15.34	-1.48	1.88	9.60	-53.00	88.67
Access to Waste Collection	695	3.29	15.84	-1.72	1.98	9.44	-70.41	94.69
Local Roads (log)	698	0.71	0.89	0.32	0.58	0.90	-7.60	11.52
Nighttime Lights (log)	698	0.34	0.21	0.23	0.33	0.45	-1.21	2.77
Panel C. Post-period Levels								
Access to Aqueduct	695	64.03	23.65	46.14	65.04	84.69	0.00	100.00
Access to Sewerage	695	46.19	25.02	27.67	43.92	63.77	0.00	100.00
Access to Waste Collection	695	53.30	25.43	33.66	51.39	72.88	0.00	109.74
Local Roads (log)	698	12.45	0.91	12.09	12.49	12.90	0.00	15.05
Nighttime Lights (log)	698	-0.49	0.76	-1.05	-0.71	-0.16	-1.38	2.71
Panel D. Pre-period Levels								
Access to Aqueduct	698	59.30	25.34	38.86	58.60	79.83	0.00	100.00
Access to Sewerage	698	42.31	25.52	21.63	40.32	60.47	0.00	100.00
Access to Waste Collection	698	50.06	25.82	29.87	48.24	68.84	0.00	107.72
Local Roads (log)	698	11.74	1.29	11.37	11.88	12.34	0.00	14.48
Nighttime Lights (log)	698	-0.82	0.88	-1.49	-1.03	-0.39	-3.75	2.82

Notes: This table reports descriptive statistics for the full experimental sample of 698 municipalities. Panel A: components are standardized to the mean and standard deviation of the change in the control group; the index is the average of the standardized components. Panels B–D: original units (percent coverage for utilities, log km for local roads, log mean luminosity for nighttime lights). Panel B reports the changes between the pre- and post-intervention periods; Panel C reports the post-intervention period; Panel D reports the pre-intervention period. Means for the control group are reported in the regression tables.

Table A2: Descriptive Statistics: Mechanisms (Procurement Practices and Corruption Investigations)

	N	Mean	SD	P25	P50	P75	Min	Max
Panel A. Standardized Outcomes (post period)								
% Value Competitive Contracts	698	0.07	0.99	-0.67	0.10	0.87	-2.15	2.13
% Value Direct Contracts	698	-0.09	0.99	-0.88	-0.27	0.60	-2.18	2.99
Corruption Investigations (log)	698	-0.10	0.91	-0.87	0.01	0.53	-0.87	3.50
Corruption Investigations (=1)	698	-0.02	1.00	-1.02	0.98	0.98	-1.02	0.98
Panel B. Original Units (post period)								
% Value Competitive Contracts	698	45.01	20.00	30.00	45.52	61.18	0.00	86.70
% Value Direct Contracts	698	40.52	19.14	25.18	37.01	53.85	0.00	100.00
Corruption Investigations (log)	698	0.61	0.72	0.00	0.69	1.10	0.00	3.43
Corruption Investigations (=1)	698	0.50	0.50	0.00	1.00	1.00	0.00	1.00
Panel C. Original Units (pre period)								
% Value Competitive Contracts	696	48.60	19.02	37.40	49.92	61.64	0.00	99.75
% Value Direct Contracts	696	36.20	17.80	23.86	33.37	46.51	0.00	100.00
Corruption Investigations (log)	698	0.50	0.63	0.00	0.00	0.69	0.00	3.00
Corruption Investigations (=1)	698	0.46	0.50	0.00	0.00	1.00	0.00	1.00

Notes: This table reports descriptive statistics for the full experimental sample of 698 municipalities. The four outcomes are those corresponding to the outcomes examined in Section 5.2: public procurement shares (% Value Competitive Contracts and % Value Direct Contracts) and the corruption-investigation measures on the intensive margin (count of cases) and the extensive margin (indicator for any case). *Panel A* reports each post-period outcome standardized to the control group mean and standard deviation used in Figure 6. *Panel B* reports the same post-period outcomes in their original units (percentages, counts, and a 0/1 indicator), as they appear in regression Tables A25 and A26. *Panel C* reports the same outcomes in their original units in the pre-intervention period (2016–2019).

Table A3: Effects on Public Service Index

	(1) Index	(2) Access to Aqueduct	(3) Access to Sewerage	(4) Access to Waste Collection	(5) Local Roads (Log)	(6) Nighttime Lights (Log)
Panel A. Pooled Treatment Effects						
[T] Any treatment	0.134***	0.065	0.184**	0.163**	0.183**	0.185**
	(0.048)	(0.080)	(0.080)	(0.082)	(0.093)	(0.074)
	[0.003]	[0.453]	[0.020]	[0.047]	[0.050]	[0.033]
Panel B. Letter vs. No Letter						
[NL] No Letter - Any Ad	0.102	-0.005	0.158	0.083	0.208	0.206*
	(0.066)	(0.099)	(0.106)	(0.101)	(0.148)	(0.121)
	[0.103]	[0.950]	[0.100]	[0.410]	[0.147]	[0.107]
[L] Letter - Any Ad	0.150***	0.101	0.196**	0.203**	0.170*	0.174**
	(0.051)	(0.087)	(0.085)	(0.090)	(0.093)	(0.075)
	[0.000]	[0.217]	[0.017]	[0.027]	[0.063]	[0.027]
Test $NL = L$, p-value	0.44	0.25	0.70	0.21	0.79	0.79
Control Mean	0.002	0.000	0.000	-0.000	0.000	-0.000
Sample Size	698	695	695	695	698	698

Notes: This table presents the effect of the intervention on the public service index and its components (expressed in z-scores). All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). Robust standard errors are shown in parentheses and randomization inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A4: Effects on Components of Public Service Index in Original Units

	(1) Access to Aqueduct	(2) Access to Sewerage	(3) Access to Waste Collection	(4) Local Roads (Log)	(5) Nighttime Lights (Log)
Panel A. Pooled Treatment Effects					
[T] Any treatment	1.166	2.866**	2.524**	0.129**	0.033**
	(1.440)	(1.256)	(1.271)	(0.065)	(0.013)
	[0.370]	[0.013]	[0.057]	[0.073]	[0.017]
Panel B. Letter vs. No Letter					
[NL] No Letter - Any Ad	-0.090	2.474	1.294	0.147	0.037*
	(1.772)	(1.661)	(1.560)	(0.104)	(0.022)
	[0.970]	[0.150]	[0.420]	[0.113]	[0.083]
[L] Letter - Any Ad	1.811	3.067**	3.155**	0.120*	0.031**
	(1.563)	(1.330)	(1.398)	(0.066)	(0.013)
	[0.243]	[0.020]	[0.030]	[0.073]	[0.023]
Test $NL = L$, p-value	0.25	0.70	0.21	0.79	0.79
Control Mean	3.959	1.984	1.575	0.647	0.315
Control Mean in Levels at Baseline	58.422	41.756	48.752	11.795	-0.848
Sample Size	695	695	695	698	698

Notes: This table presents the effect of the intervention on components of the public service index measured in their original units. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). Robust standard errors are shown in parentheses and randomization inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A5: Effect on Public Service Index by Urban vs. Rural

	(1) Index	(2) Index Urban	(3) Index Rural
Panel A. Pooled Treatment Effects			
[T] Any treatment	0.134***	0.142***	0.088*
	(0.048)	(0.048)	(0.045)
	[0.007]	[0.013]	[0.033]
Panel B. Letter - No Letter			
[NL] No Letter - Any Ad	0.102	0.149**	0.085
	(0.066)	(0.064)	(0.060)
	[0.160]	[0.017]	[0.140]
[L] Letter - Any Ad	0.150***	0.138***	0.089*
	(0.051)	(0.052)	(0.048)
	[0.003]	[0.003]	[0.067]
Test $NL = L$, p-value	0.44	0.86	0.95
Control Mean	0.002	0.002	0.001
Sample Size	698	698	698

Notes: This table presents the effects of the intervention on the public service index, by area of coverage. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). Robust standard errors are shown in parentheses, and randomization inference p-values are shown in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A6: Effects on Access to Services Using Household Surveys

	(1) Index	(2) PCA_1	(3) Sewerage	(4) Waste Collection	(5) Acqueduct
Panel A. Pooled Treatment					
$[T]$ Any treatment	0.080***	0.079***	0.008	0.043***	0.042**
	(0.027)	(0.027)	(0.015)	(0.014)	(0.019)
	[0.007]	[0.007]	[0.657]	[0.007]	[0.017]
Panel B. Letter vs. No Letter					
$[NL]$ No Letter - Any Ad	0.043	0.042	-0.008	0.021	0.035
	(0.037)	(0.036)	(0.022)	(0.020)	(0.024)
	[0.207]	[0.220]	[0.657]	[0.213]	[0.170]
$[L]$ Letter - Any Ad	0.099***	0.098***	0.017	0.054***	0.046**
	(0.030)	(0.029)	(0.017)	(0.016)	(0.019)
	[0.003]	[0.003]	[0.387]	[0.000]	[0.027]
Test $NL = L$, p-value	0.13	0.13	0.28	0.12	0.58
Control Mean	0.00	0.00	0.40	0.54	0.69
Sample Size	103040	103040	103040	103040	103040
N. Municipalities	469	469	469	469	469

Note: This table presents the effects of the intervention on public services measured in the 2022-2023 ECV household surveys. The outcome in column (1) corresponds to an index of the standardized values for access to sewerage, waste collection, and aqueduct. In column (2) it corresponds to the first principal component from a principal components analysis (PCA) of the same three variables. Columns (3)–(5) correspond to responses to the question "Which of the following utilities does the household have?" (2) Sewerage: 1 = "Yes"; (3) Waste collection: 1 = "Yes"; (4) Water: 1 = "Yes". All estimates use sampling weights provided by DANE. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). An indicator for whether a household is rural and an indicator for the year of the survey are included in the set of potential covariates. Robust standard errors are shown in parentheses, and randomization inference p-values are shown in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A7: Effects on GDP and Components of Nighttime Lights

	(1)	(2)	(3)
	Log GDP	Residual Nighttime Lights (Log)	Predicted Nighttime Lights (Log) By GDP
Panel A. Pooled Treatment Effects			
[T] Any treatment	-0.015	0.033***	-0.006
	(0.016)	(0.013)	(0.006)
	[0.317]	[0.013]	[0.347]
Panel B. Letter vs. No Letter			
[NL] No Letter - Any Ad	-0.027	0.037**	-0.011
	(0.021)	(0.018)	(0.008)
	[0.183]	[0.030]	[0.170]
[L] Letter - Any Ad	-0.008	0.031**	-0.003
	(0.017)	(0.013)	(0.007)
	[0.637]	[0.023]	[0.677]
Test $NL = L$, p-value	0.30	0.76	0.30
Control Mean	6.01	-0.06	-0.47
Sample Size	698	698	698

Notes: This table presents the effects of the intervention on the GDP of municipalities, as well as the components of the night time lights predicted and non-predicted by GDP. The outcome in column (1) is the (log) average GDP in 2022 and 2023. In column (2) it is the (log) nighttime light average in 2022 and 2023 once it has been residualized in a linear regression from the (log) average GDP in 2022 and 2023. In column (3) it is the (log) nighttime light average in 2022 and 2023 predicted in a linear regression using the (log) average GDP in 2022 and 2023 as an explanatory variable. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). Robust standard errors are shown in parentheses, and randomization inference p-values are shown in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A8: Effects on Public Service Index Measured in Levels

	(1) Index	(2) Access to Aqueduct	(3) Access to Sewerage	(4) Access to Waste Collection	(5) Local Roads (Log)	(6) Nighttime Lights (Log)
Panel A. Pooled Treatment Effects						
[T] Any treatment	0.071***	0.044	0.101**	0.090**	0.081	0.041***
	(0.025)	(0.047)	(0.044)	(0.041)	(0.060)	(0.014)
	[0.000]	[0.343]	[0.010]	[0.033]	[0.080]	[0.007]
Panel B. Letter vs. No Letter						
[NL] No Letter - Any Ad	0.046	0.005	0.078	0.045	0.070	0.043*
	(0.035)	(0.061)	(0.059)	(0.054)	(0.070)	(0.022)
	[0.150]	[0.957]	[0.180]	[0.407]	[0.373]	[0.040]
[L] Letter - Any Ad	0.083***	0.065	0.112**	0.114**	0.086	0.041***
	(0.026)	(0.051)	(0.047)	(0.045)	(0.060)	(0.015)
	[0.003]	[0.197]	[0.020]	[0.007]	[0.090]	[0.010]
Test $NL = L$, p-value	0.26	0.29	0.54	0.19	0.71	0.94
Control Mean	0.281	0.156	0.080	0.061	0.685	0.403
Sample Size	698	695	695	695	698	698

Notes: This table presents the effect of the intervention on on a version of the public service index and its components measured in levels in 2022-2023 (expressed as z-scores). All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). A baseline version of the index and its components measured in 2018-2019, as explained in the main text, is included in the set of possible controls. Robust standard errors are shown in parentheses, and randomization inference p-values are shown in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A9: Estimates of the Full Set of Treatment Arms on Public Service Index

	(1) Index	(2) Access to Aqueduct	(3) Access to Sewerage	(4) Access to Waste Collection	(5) Local Roads (Log)	(6) Nighttime Lights (Log)
Info + No Letter	0.270**	0.275**	0.443**	0.308**	0.224	0.347
	(0.112)	(0.137)	(0.182)	(0.137)	(0.237)	(0.268)
	[0.000]	[0.057]	[0.000]	[0.030]	[0.210]	[0.047]
CtA + No Letter	0.010	-0.053	0.097	0.046	0.073	0.006
	(0.092)	(0.132)	(0.148)	(0.149)	(0.110)	(0.195)
	[0.917]	[0.737]	[0.533]	[0.787]	[0.567]	[0.990]
Info + CtA + No Letter	0.026	-0.243	-0.070	-0.108	0.327	0.246**
	(0.094)	(0.152)	(0.133)	(0.149)	(0.313)	(0.107)
	[0.803]	[0.147]	[0.613]	[0.517]	[0.120]	[0.087]
Info + Partial K. Letter	0.097	0.123	0.179	0.227	0.161	-0.063
	(0.109)	(0.186)	(0.145)	(0.206)	(0.164)	(0.198)
	[0.313]	[0.457]	[0.253]	[0.127]	[0.287]	[0.707]
CtA + Partial K. Letter	0.087	0.067	-0.007	0.049	0.180	0.260**
	(0.077)	(0.138)	(0.125)	(0.134)	(0.120)	(0.113)
	[0.313]	[0.623]	[0.960]	[0.763]	[0.187]	[0.057]
Info + CtA + Partial K. Letter	0.187**	0.166	0.191	0.181*	0.253	0.128
	(0.077)	(0.130)	(0.117)	(0.109)	(0.244)	(0.117)
	[0.037]	[0.277]	[0.210]	[0.220]	[0.140]	[0.353]
Info + Full K. Letter	0.070	0.072	0.073	0.148	-0.024	0.139
	(0.084)	(0.151)	(0.125)	(0.134)	(0.097)	(0.120)
	[0.457]	[0.580]	[0.637]	[0.353]	[0.847]	[0.253]
CtA + Full K. Letter	0.232**	0.018	0.202	0.254	0.415*	0.378***
	(0.110)	(0.173)	(0.191)	(0.214)	(0.227)	(0.118)
	[0.017]	[0.923]	[0.200]	[0.137]	[0.007]	[0.003]
Info + CtA + Full K. Letter	0.223***	0.154	0.516***	0.349**	0.039	0.155
	(0.068)	(0.128)	(0.148)	(0.148)	(0.103)	(0.109)
	[0.010]	[0.317]	[0.000]	[0.027]	[0.723]	[0.200]
Test: Any treatment is equal, p-value	0.21	0.21	0.03	0.29	0.43	0.42
Test: Info treatments are equal, p-value	0.28	0.50	0.18	0.63	0.34	0.44
Test: CtA treatments are equal, p-value	0.25	0.77	0.57	0.65	0.31	0.22
Test: Info + CtA treatments are equal, p-value	0.15	0.04	0.00	0.05	0.48	0.64
Test: No Letter treatments are equal, p-value	0.11	0.01	0.04	0.06	0.63	0.45
Test: Letter treatments are equal, p-value	0.42	0.97	0.06	0.68	0.23	0.26
Test: Partial K. Letter treatments are equal, p-value	0.53	0.83	0.33	0.60	0.94	0.28
Test: Full K. Letter treatments are equal, p-value	0.22	0.76	0.03	0.52	0.15	0.19
Control Mean	0.00	0.00	0.00	-0.00	0.00	-0.00
Sample Size	698	695	695	695	698	698

Notes: This table presents the effects of the full set of treatment arms on the public service index and its components (expressed as z-scores). The p-values of different tests of equality of estimates are displayed in the bottom of the table. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). Robust standard errors are shown in parentheses, and randomization inference p-values are shown in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A10: Effects on Public Service Index Without Controls

	(1) Index	(2) Access to Aqueduct	(3) Access to Sewerage	(4) Access to Waste Collection	(5) Local Roads (Log)	(6) Nighttime Lights (Log)
Panel A. Pooled Treatment Effects						
[T] Any treatment	0.144***	0.063	0.188**	0.163**	0.131	0.182**
	(0.048)	(0.080)	(0.081)	(0.082)	(0.091)	(0.083)
	[0.000]	[0.423]	[0.017]	[0.043]	[0.163]	[0.017]
Panel B. Letter vs. No Letter						
[NL] No Letter - Any Ad	0.125*	-0.012	0.164	0.080	0.169	0.235*
	(0.067)	(0.100)	(0.107)	(0.101)	(0.147)	(0.129)
	[0.040]	[0.903]	[0.137]	[0.397]	[0.237]	[0.067]
[L] Letter - Any Ad	0.154***	0.101	0.200**	0.205**	0.112	0.156*
	(0.051)	(0.086)	(0.085)	(0.090)	(0.094)	(0.084)
	[0.000]	[0.207]	[0.003]	[0.023]	[0.233]	[0.067]
Test $NL = L$, p-value	0.65	0.22	0.72	0.20	0.70	0.53
Control Mean	0.002	0.000	0.000	-0.000	0.000	-0.000
Sample Size	698	695	695	695	698	698

Notes: This table presents the effect of the intervention on the public service index and its components (expressed as z-scores) without controls. Only strata fixed effects are used as controls. Robust standard errors are shown in parentheses and randomization inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A11: Effects on Elected Mayors' Cleanliness

	(1) Above Average Cleanliness Index (=1)	(2) Above Average Cleanliness Index, Residualized Popularity (=1)	(3) Below Average Will Perform Irregularities (% Responses)	(4) Will Perform Irregularities (% Responses)
Panel A. Pooled Treatment Effect				
[T] Any treatment	0.071*	0.099**	0.077*	-0.037
	(0.041)	(0.043)	(0.041)	(0.023)
	[0.087]	[0.020]	[0.067]	[0.100]
Panel B. Subtreatments by Letter - No Letter				
[NL] No Letter - Any Ad	0.053	0.085	0.069	-0.015
	(0.053)	(0.055)	(0.053)	(0.029)
	[0.333]	[0.143]	[0.240]	[0.623]
[L] Letter - Any Ad	0.080*	0.105**	0.081*	-0.047*
	(0.044)	(0.046)	(0.044)	(0.025)
	[0.087]	[0.037]	[0.083]	[0.060]
Test $NL = L$, p-value	0.61	0.69	0.82	0.24
Control Mean	0.31	0.36	0.31	0.62
Sample Size	595	595	592	591

Notes: This table presents the effect of the intervention on different measures of elected mayors' cleanliness. The sample corresponds to the 595 municipalities for which the standardized cleanliness index used in column (1) is defined; that is, municipalities where the two candidates with the highest vote shares had responses in the pre-treatment 2019 citizen survey and received different standardized cleanliness scores, allowing a "cleaner candidate" to be identified. The outcome in column (1) is an indicator for whether the elected mayor's cleanliness score is above the municipal average. Column (2) uses the same indicator, but the cleanliness index is first residualized with respect to candidate popularity, measured by the pre-treatment survey expected-winner score. Column (3) reports an indicator for whether the share of pre-treatment survey respondents in the municipality who expected the elected candidate to commit irregularities (Z_{mc}, described in the main text) is below the municipal average. Column (4) reports the elected mayor's value of Z_{mc}. Sample sizes are slightly smaller in columns (3) and (4) because the raw Z_{mc} measure is (i) tied across the two leading candidates in 3 municipalities and (ii) missing for the elected mayor in 4 municipalities because the survey did not ask about that candidate. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). Robust standard errors are reported in parentheses, and randomization-inference p-values are reported in square brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A12: Descriptive Statistics: Decomposition

	N	Mean	SD	P25	P50	P75	Min	Max
Panel A. Cleaner Candidate Won (=1)								
Cleanliness Index	595	0.365	0.482	0.000	0.000	1.000	0.000	1.000
Cleanliness Index, residualized on winner pop.	595	0.425	0.495	0.000	0.000	1.000	0.000	1.000
$Z_{c,m}$	592	0.370	0.483	0.000	0.000	1.000	0.000	1.000
Panel B. Running Variable: Margin of Victory								
Cleanliness Index	595	-7.997	17.070	-19.971	-6.889	4.983	-51.393	34.015
Cleanliness Index, residualized on winner pop.	595	-5.223	16.892	-16.905	-3.087	6.586	-51.393	43.563
Panel C. Additional Cleanliness Variables								
% respondents expecting irregularities (municipality mean)	591	0.594	0.262	0.429	0.600	0.786	0.000	1.000
Past misconduct indicator (municipality share)	47	0.553	0.503	0.000	1.000	1.000	0.000	1.000

Notes: This table reports descriptive statistics for the sample restricted to the 595 municipalities for which the standardized *Cleanliness Index* is defined; that is, municipalities where the two candidates with the highest vote shares had responses in the pre-treatment 2019 citizen survey and received different standardized cleanliness scores, allowing a "cleaner" candidate to be identified. The first three rows are indicator variables equal to one if the elected mayor was the cleaner of the two top vote-getters under three alternative cleanliness measures: the standardized *Cleanliness Index*; the same index residualized with respect to the pre-treatment survey expected-winner score; and the raw survey measure $Z_{c,m}$. Rows 4 and 5 report the corresponding running variables, defined as the vote-share margin between the cleaner and less-clean top candidate. The running variable based on the standardized *Cleanliness Index* is the one used in the RD analysis. The last two rows report the underlying candidate-level measures aggregated to the municipality level: the share of respondents who expected the elected candidate to commit irregularities and an indicator for whether the candidate had a record of past misconduct.

Table A13: RD Robustness to Experimental Treatment Arms

	(1) Original Estimate (1)	(2) Controlling For Treatment Dummy (2)	(3) Only Control Group (3)	(4) Only Treated Group (4)
RD Estimate	0.412^{**}	0.522^{**}	0.246	0.770^{***}
	(0.204)	(0.207)	(0.303)	(0.265)
Bandwidth selection				
Bandwidth	8.38	7.12	9.14	6.55
Conventional p-value	0.04	0.01	0.44	0.00
Bias-Robust p-value	0.04	0.01	0.42	0.00
Sample Size	209	187	72	116

Notes: This table presents the robustness of the RD estimates to the different experimental treatment arms. Column (1) presents the main estimate from Table 1. Column (2) controls for a treatment indicator. Columns (3) and (4) display the estimated effect only using the control and the pooled treatment group samples, respectively. Standard errors in parenthesis are the ones estimated by the robust bias-corrected method proposed by Calonico et al. (2014); conventional and bias-robust p-values are shown in the lower part of the figure; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A14: RD Estimates - Residualizing Popularity

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. First Order Polynomial						
RD Estimate	0.407**	0.465*	0.376**	0.310**	0.333	0.292**
	(0.199)	(0.262)	(0.180)	(0.155)	(0.338)	(0.139)
Bandwidth selection	Optimal	0.7 × Optimal	1.3 × Optimal	Optimal	0.7 × Optimal	1.3 × Optimal
Bandwidth	9.04	6.33	11.75	8.24	5.77	10.71
Controls				X	X	X
Conventional p-value	0.03	0.03	0.02	0.04	0.04	0.04
Bias-Robust p-value	0.04	0.08	0.04	0.05	0.32	0.04
Sample Size	252	187	311	236	175	293
Panel B. Second Order Polynomial						
RD Estimate	0.421*	0.488*	0.427*	0.384**	0.415	0.339**
	(0.226)	(0.269)	(0.223)	(0.189)	(0.283)	(0.168)
Bandwidth selection	Optimal	0.7 × Optimal	1.3 × Optimal	Optimal	0.7 × Optimal	1.3 × Optimal
Bandwidth	16.14	11.30	20.99	14.35	10.04	18.65
Controls				X	X	X
Conventional p-value	0.04	0.04	0.02	0.03	0.03	0.04
Bias-Robust p-value	0.06	0.07	0.06	0.04	0.14	0.04
Sample Size	406	306	472	362	279	436

Notes: This table presents alternative RD estimates when candidate popularity is residualized from the cleanliness measure as described in the text. The specifications in columns (4)-(6) include the controls displayed in Figure A8. Standard errors in parenthesis are the ones estimated by the robust bias-corrected method proposed by Calonico et al. (2014); conventional and bias-robust p-values are shown in the lower part of the figure; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A15: Selection Effect Estimates

	(1)	(2)	(3)	(4)
	Baseline Index		Residualized Popularity	
Estimate	21.823 %	21.211 %	29.968%	24.930%
	[0.070]	[0.057]	[0.036]	[0.060]
RD Controls		X		X

Notes: This table presents the estimates of the selection effect as a percentage of the average treatment effect. Columns (1) and (2) display the estimates using the baseline index of cleanliness, while columns (3) and (4) display the estimates obtained when residualizing the popularity of candidates from the measure of cleanliness, as explained in the main text. RD controls are used to produce the estimates in the indicated columns. Bootstrapped p-values are shown in square brackets.

Table A16: Correlation Matrix of Candidate Characteristics

	(1) Log Age	(2) Female (=1)	(3) Undergraduate or More (=1)	(4) Business Owner (=1)	(5) Any Public Sector Experience (=1)	(6) Any Private Sector Experience (=1)	(7) Any Past Pol. Post (=1)
ρ - Above Average Clean. Index (=1)	-0.106***	0.032	-0.036	-0.023	-0.051	0.038	-0.085**
	(0.009)	(0.444)	(0.470)	(0.579)	(0.325)	(0.451)	(0.036)
ρ - Will Perform Irreg. (% Responses)	0.051	-0.010	0.004	-0.008	-0.008	0.004	0.042
	(0.208)	(0.799)	(0.932)	(0.845)	(0.868)	(0.928)	(0.294)

	(8) N. Political Posts	(9) Incumbent (=1)	(10) Incumbent Loose Def. (=1)	(11) Coalition of Parties (=1)	(12) Independent Candidate (=1)	(13) Historical Party (=1)	(14) Institutional Left Party (=1)
ρ - Above Average Clean. Index (=1)	-0.048	-0.008	-0.001	0.035	-0.055	0.065	0.045
	(0.240)	(0.840)	(0.986)	(0.398)	(0.137)	(0.123)	(0.295)
ρ - Will Perform Irreg. (% Responses)	0.046	-0.044	-0.016	-0.006	0.045	-0.091**	-0.012
	(0.258)	(0.272)	(0.691)	(0.890)	(0.261)	(0.023)	(0.761)

	(1) Share Private Donations	(2) Share Self-Funded	(3) Share Party-Funded	(4) Log Total Campaign Income
ρ - Above Average Clean. Index (=1)	-0.020	0.045	-0.015	-0.109***
	(0.628)	(0.266)	(0.700)	(0.003)
ρ - Will Perform Irreg. (% Responses)	0.054	-0.061	0.058	0.088**
	(0.183)	(0.128)	(0.153)	(0.029)

Notes: This table presents the correlations of different candidate characteristics with the dichotomous measure of candidate cleanliness, as well as the continuous measure of the percentage of survey respondents stating that the candidate will conduct irregularities. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A17: Effects on Characteristics of Elected Mayor: Socio-demographics and Experience

	(1) Log Age	(2) Female (=1)	(3) Undergraduate or More (=1)	(4) Business Owner (=1)	(5) Any Public Sector Experience (=1)	(6) Any Private Sector Experience (=1)
Pooled Treatment Effects						
[T] Any treatment	-0.019	-0.010	0.076	-0.032	-0.039	0.003
	(0.016)	(0.026)	(0.049)	(0.040)	(0.030)	(0.050)
	[0.197]	[0.670]	[0.113]	[0.433]	[0.237]	[0.923]
Letter vs. No Letter						
[NL] No Letter - Any Ad	-0.025	-0.012	0.019	-0.039	-0.026	0.008
	(0.021)	(0.033)	(0.062)	(0.051)	(0.039)	(0.063)
	[0.210]	[0.740]	[0.807]	[0.450]	[0.557]	[0.910]
[L] Letter - Any Ad	-0.016	-0.009	0.106**	-0.028	-0.046	0.001
	(0.017)	(0.029)	(0.053)	(0.043)	(0.034)	(0.054)
	[0.307]	[0.747]	[0.040]	[0.493]	[0.213]	[0.993]
Test $NL = L$, p-value	0.64	0.93	0.13	0.81	0.61	0.90
Control Mean	3.78	0.12	0.47	0.43	0.91	0.47
Sample Size	697	697	462	695	462	462

Notes: This table presents the effects of the intervention on the characteristics of elected Mayors related to their socio-demographic profile and previous professional experience. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). Robust standard errors are reported in parentheses, and randomization-inference p-values are reported in square brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A18: Effects on Characteristics of Elected Mayor: Political and Party Profile

	(1) Any Past Pol. Post (=1)	(2) N. Political Posts	(3) Incumbent (=1)	(4) Incumbent Loose Def. (=1)	(5) Coalition of Parties (=1)	(6) Independent Candidate (=1)	(7) Historical Party (=1)	(8) Institutional Left Party (=1)
Pooled Treatment Effects								
[T] Any treatment	-0.067	-0.039	0.024	0.057	0.007	-0.021	-0.020	0.018
	(0.076)	(0.040)	(0.020)	(0.037)	(0.038)	(0.014)	(0.032)	(0.020)
	[0.323]	[0.320]	[0.223]	[0.133]	[0.890]	[0.093]	[0.507]	[0.410]
Letter vs. No Letter								
[NL] No Letter - Any Ad	-0.070	-0.058	0.005	0.007	-0.018	-0.023	-0.050	0.017
	(0.098)	(0.051)	(0.024)	(0.046)	(0.049)	(0.015)	(0.039)	(0.026)
	[0.470]	[0.280]	[0.860]	[0.907]	[0.750]	[0.267]	[0.210]	[0.487]
[L] Letter - Any Ad	-0.065	-0.030	0.034	0.083**	0.020	-0.020	-0.004	0.019
	(0.081)	(0.044)	(0.023)	(0.040)	(0.041)	(0.014)	(0.035)	(0.022)
	[0.427]	[0.533]	[0.113]	[0.040]	[0.660]	[0.170]	[0.920]	[0.447]
Test $NL = L$, p-value	0.96	0.55	0.25	0.09	0.41	0.75	0.21	0.93
Control Mean	0.72	0.46	0.06	0.27	0.43	0.04	0.21	0.06
Sample Size	698	698	698	698	698	698	698	698

Notes: This table presents the effects of the intervention on the characteristics of elected Mayors related to their political profile and the characteristics of the electoral platform used to run for office. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). Robust standard errors are reported in parentheses, and randomization-inference p-values are reported in square brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A19: Effects on Characteristics of Elected Mayor: Campaign Finance Profile

	(1) Share Private Donations	(2) Share Self-Funded	(3) Share Party-Funded	(4) Log Total Campaign Income
Pooled Treatment Effects				
[T] Any treatment	-0.018	0.026	-0.004	0.038
	(0.023)	(0.024)	(0.003)	(0.094)
	[0.373]	[0.237]	[0.143]	[0.643]
Letter vs. No Letter				
[NL] No Letter - Any Ad	-0.027	0.035	-0.005	0.093
	(0.028)	(0.030)	(0.004)	(0.101)
	[0.327]	[0.227]	[0.257]	[0.507]
[L] Letter - Any Ad	-0.013	0.021	-0.004	0.010
	(0.024)	(0.026)	(0.003)	(0.096)
	[0.580]	[0.383]	[0.300]	[0.937]
Test $NL = L$, p-value	0.58	0.62	0.70	0.15
Control Mean	0.22	0.75	0.01	17.65
Sample Size	697	697	697	698

Notes: This table presents the effects of the intervention on the characteristics of elected Mayors related to their campaign funding. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). Robust standard errors are reported in parentheses, and randomization-inference p-values are reported in square brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A20: Effects on Reports to MOE in 2023 Elections

	(1) Reports per 1000 citizens	(2) Asinh Transform Count	(3) Any report = 1
Panel A. Pooled Treatment Effects			
[T] Any treatment	0.025**	0.121*	0.043
	(0.011)	(0.072)	(0.035)
	[0.040]	[0.097]	[0.223]
Panel B. Letter vs. No Letter			
[NL] No Letter - Any Ad	0.014	0.022	-0.020
	(0.013)	(0.095)	(0.046)
	[0.280]	[0.793]	[0.707]
[L] Letter - Any Ad	0.030**	0.172**	0.076**
	(0.012)	(0.077)	(0.037)
	[0.013]	[0.030]	[0.020]
Test $NL = L$, p-value	0.24	0.09	0.02
Control Mean	0.099	1.159	0.707
Sample Size	698	698	698

Notes: This table presents the effect of the intervention on citizen reports to the MOE around the 2023 elections. Column (1) shows the number of reports per 1,000 citizens, column (2) shows the inverse hyperbolic sine transformation of the total number of reports in 2023 (which accounts for the large number of zeros), and column (3) shows an indicator that takes the value of 1 if the latter variable is positive. All specifications include the covariates selected using the 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: Effects on Protests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Only Municipal Protests			
		By Geography		By Reason		By Adversary	
	All Protests (per 10K pop.)	Municipality	Rest	Policies and Public SS	Rest	Executive Govmt.	Rest
Panel A. Pooled Treatment Effects							
[T] Any treatment	-0.076	-0.049	0.000	-0.018	-0.041	-0.019	-0.039
	(0.123)	(0.079)	(0.083)	(0.033)	(0.063)	(0.049)	(0.052)
	[0.530]	[0.490]	[1.000]	[0.550]	[0.500]	[0.667]	[0.437]
Panel B. Letter vs. No Letter							
[NL] No Letter - Any Ad	0.094	0.070	0.032	0.049	0.008	0.049	0.012
	(0.162)	(0.101)	(0.109)	(0.048)	(0.077)	(0.064)	(0.063)
	[0.537]	[0.450]	[0.787]	[0.253]	[0.940]	[0.427]	[0.843]
[L] Letter - Any Ad	-0.165	-0.111	-0.016	-0.052	-0.067	-0.054	-0.065
	(0.131)	(0.084)	(0.089)	(0.033)	(0.069)	(0.053)	(0.057)
	[0.170]	[0.207]	[0.827]	[0.120]	[0.327]	[0.290]	[0.277]
Test $NL = L$, p-value	0.09	0.05	0.64	0.02	0.30	0.09	0.20
Control Mean	1.544	0.683	0.861	0.199	0.483	0.332	0.350
Sample Size	698	698	698	698	698	698	698

Notes: This table presents the effects of the intervention on protests in 2020-2023. The outcome in column (1) is the number of any protests occurring per 10K inhabitants. In column (2) it is the number of protests per 10K inhabitants occurring where the geographic coverage is the municipal or submunicipal levels, while in column (3) it includes other levels of coverage. In column (4) it is the subset of municipal or submunicipal protests where the main reason for the protests concerns public policies and social security, while in column (5) it groups other reasons. In column (6) it is the subset of municipal or submunicipal protests where the main adversary is categorized to be the executive government, while in column (7) it groups other adversaries. All specifications include covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). A baseline version of each variable measured in 2016-2019 is included in the set of possible controls. Robust standard errors are reported in parentheses, and randomization inference p-values are reported in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A22: Effects on Grass-root Citizen Associations

	(1) Number of All Associations	(2) Number of JACs	(3) Members of All Associations (% population)	(4) Members JACs (% population)
Panel A. Pooled Treatment Effects				
[T] Any treatment	0.346	0.496	0.785	0.578
	(2.139)	(2.059)	(0.733)	(0.711)
	[0.840]	[0.760]	[0.297]	[0.447]
Panel B. Letter vs. No Letter				
[NL] No Letter - Any Ad	-0.258	-0.055	0.732	0.467
	(2.578)	(2.493)	(0.930)	(0.902)
	[0.933]	[0.990]	[0.447]	[0.570]
[L] Letter - Any Ad	0.660	0.781	0.812	0.636
	(2.364)	(2.277)	(0.820)	(0.788)
	[0.823]	[0.730]	[0.370]	[0.450]
Test $NL = L$, p-value	0.71	0.73	0.93	0.85
Control Mean	37.471	36.044	12.463	12.155
Sample Size	698	698	698	698

Notes: This table presents the effects of the intervention on grassroots associations in 2020-2023. The outcome in column (1) is the number of all grassroots associations, while in column (2) it is the subset of *Juntas de Acción Comunal* associations. In column (3) it is the members of any association expressed as the percentage of the population, while in column (4) it is the members of *Juntas de Acción Comunal* as a percentage of the population. All specifications include covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). A baseline version of each variable measured in 2016-2019 is included in the set of possible controls. Robust standard errors are reported in parentheses, and randomization inference p-values are reported in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A23: Effects on Citizen Writs (*Tutelas*)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		By Right Being Claimed (per 1K pop.)					
	All Writs (per 1K pop.)	Public Works	Social Services	Health	FOIA	Justice	Labor
Panel A. Pooled Treatment Effects							
[T] Any treatment	-0.019	0.004	-0.024**	0.006	0.008	-0.020	-0.001
	(0.102)	(0.003)	(0.011)	(0.041)	(0.038)	(0.013)	(0.003)
	[0.857]	[0.327]	[0.013]	[0.883]	[0.833]	[0.217]	[0.640]
Panel B. Letter vs. No Letter							
[NL] No Letter - Any Ad	0.034	0.010	-0.018	0.048	0.022	-0.027*	-0.001
	(0.131)	(0.007)	(0.014)	(0.043)	(0.045)	(0.015)	(0.003)
	[0.843]	[0.107]	[0.203]	[0.397]	[0.637]	[0.100]	[0.760]
[L] Letter - Any Ad	-0.046	0.002	-0.027**	-0.016	0.001	-0.016	-0.001
	(0.108)	(0.003)	(0.011)	(0.046)	(0.044)	(0.016)	(0.003)
	[0.653]	[0.603]	[0.013]	[0.730]	[0.990]	[0.357]	[0.670]
Test $NL = L$, p-value	0.50	0.25	0.42	0.09	0.68	0.59	0.91
Control Mean	2.24	0.03	0.16	0.74	0.62	0.23	0.05
Sample Size	698	698	698	698	698	698	698

Notes: This table presents the effects of the intervention on *tutelas* in 2020-2023. The outcome in column (1) is the total number of writs per 1K inhabitants. The outcomes in columns (2)-(7) correspond to the number of writs per 1K inhabitants by type of right being claimed. All specifications include covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). A baseline version of each variable measured in 2019 is included in the set of possible controls. Robust standard errors are reported in parentheses, and randomization inference p-values are reported in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A24: Effects on Active Social Media Users

	(1)	(2)	(3)	(4)
	Meta Users (Log)	Meta Users (% pop.)	Meta Users Interested in Political and Social Issues (Log)	Meta Users Interested in Political and Social Issues (% pop.)
Panel A. Pooled Treatment Effects				
[T] Any treatment	-0.033	-0.615	-0.030	0.077
	(0.036)	(2.421)	(0.033)	(1.015)
	[0.387]	[0.783]	[0.333]	[0.950]
Panel B. Letter vs. No Letter				
[NL] No Letter - Any Ad	-0.014	-1.431	-0.001	-0.169
	(0.042)	(2.717)	(0.039)	(1.158)
	[0.737]	[0.600]	[0.990]	[0.887]
[L] Letter - Any Ad	-0.043	-0.199	-0.045	0.202
	(0.040)	(2.737)	(0.036)	(1.142)
	[0.237]	[0.937]	[0.210]	[0.887]
Test $NL = L$, p-value	0.47	0.64	0.24	0.74
Control Mean	9.589	77.256	8.690	31.985
Sample Size	698	698	698	698

Notes: This table presents the effects of the intervention on social media users, measured in 2026. The outcome in column (1) is the (log) number of Meta users, while in column (2) it is the number of such users as a percentage of the population. In column (3) is the (log) number of Meta users which are tagged as having interests in political or social issues, as explained in the main text, while in column (4) it is the number of such users as a percentage of the population. All specifications include covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). Robust standard errors are reported in parentheses, and randomization inference p-values are reported in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A25: Effects on Procurement Outcomes

	(1) Log Value Total Contracts 2020-2023	(2) Log Value Competitive Contracts 2020-2023	(3) % Value Competitive Contracts 2020-2023	(4) Log Value Direct Contracts 2020-2023	(5) % Value Direct Contracts 2020-2023
Panel A. Pooled Treatment Effects					
[T] Any treatment	0.053	0.180**	2.264*	-0.036	-3.119**
	(0.056)	(0.074)	(1.292)	(0.055)	(1.314)
	[0.337]	[0.023]	[0.103]	[0.477]	[0.023]
Panel B. Letter vs. No Letter					
[NL] No Letter - Any Ad	0.062	0.186**	1.062	0.010	-2.422
	(0.065)	(0.087)	(1.657)	(0.063)	(1.617)
	[0.327]	[0.057]	[0.543]	[0.843]	[0.127]
[L] Letter - Any Ad	0.049	0.177**	2.886**	-0.060	-3.482**
	(0.063)	(0.081)	(1.400)	(0.063)	(1.441)
	[0.450]	[0.043]	[0.040]	[0.337]	[0.013]
Test $NL = L$, p-value	0.83	0.91	0.25	0.27	0.49
Control Mean	24.60	23.64	43.59	23.64	42.18
Sample Size	698	692	696	697	696

Notes: This table presents the effects of the intervention on procurement outcomes. The outcome in column (1) is the (log) total value of contracts in 2020-2023. In columns (2) and (4) it is the (log) value of competitive and direct contracts, respectively. In columns (3) and (5) it is the value of competitive and direct contracts expressed as a percentage of the total value of all contracts, respectively. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). A baseline version of each variable measured in 2016-2019 is included in the set of possible controls. Robust standard errors are shown in parentheses, and randomization inference p-values are shown in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A26: Effects on Judicial Processes Against Mayors' Offices

	(1)	(2)	(3)	(4)	(5)	(6)
	Intensive Margin (Log)			Extensive Margin (=1)		
	All Processes	Corruption	Non-corruption	All Processes	Corruption	Non-corruption
Panel A. Pooled Treatment Effects						
[T] Any treatment	-0.068	-0.102*	0.011	0.028	-0.014	0.019
	(0.058)	(0.053)	(0.034)	(0.037)	(0.038)	(0.034)
	[0.210]	[0.020]	[0.777]	[0.437]	[0.707]	[0.647]
Panel B. Letter vs. No Letter						
[NL] No Letter - Any Ad	-0.099	-0.116*	-0.011	0.019	-0.009	-0.002
	(0.069)	(0.063)	(0.042)	(0.047)	(0.048)	(0.042)
	[0.167]	[0.077]	[0.807]	[0.677]	[0.830]	[1.000]
[L] Letter - Any Ad	-0.051	-0.095*	0.022	0.033	-0.016	0.030
	(0.063)	(0.057)	(0.037)	(0.040)	(0.041)	(0.037)
	[0.397]	[0.093]	[0.570]	[0.460]	[0.690]	[0.497]
Test $NL = L$, p-value	0.45	0.72	0.42	0.76	0.87	0.42
Control Mean	0.761	0.685	0.203	0.538	0.511	0.213
Sample Size	698	698	698	698	698	698

Notes: This table presents the effects of the intervention on investigations conducted by the judiciary to Mayors' Offices . The outcome in column (1) is the $\log(x+1)$ transformation of the number of investigations in 2020-2023, while in column (4) it is an indicator for any investigation. In column (2) is the $\log(x+1)$ transformation of the number of investigations related to corruption that took place in 2020-2023, while in column (5) it is an indicator for such investigations. In column (3) it is the $\log(x+1)$ transformation of the number of investigations not related to corruption that took place in 2020-2023, while in column (6) it is an indicator for such investigations. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). A baseline version of each variable measured in 2016-2019 is included in the set of possible controls. Robust standard errors are shown in parentheses, and randomization inference p-values are shown in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A27: RD Estimates on Mayoral Actions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	% Value Competitive Contracts		% Value Direct Contracts		Corruption Investigations (Log)		Corruption Investigations (=1)	
Panel A. First Order Polynomial								
RD Estimate	-9.976*	-11.634**	7.234	8.857*	-0.242	-0.401**	-0.360**	-0.444***
	(5.101)	(4.662)	(5.480)	(4.853)	(0.198)	(0.182)	(0.152)	(0.137)
Bandwidth selection	Optimal	Optimal	Optimal	Optimal	Optimal	Optimal	Optimal	Optimal
Bandwidth	10.92	8.52	10.66	8.73	10.61	8.12	7.17	6.43
Controls		X		X		X		X
Conventional p-value	0.04	0.01	0.14	0.05	0.22	0.03	0.02	0.00
Bias-Robust p-value	0.05	0.01	0.19	0.07	0.22	0.03	0.02	0.00
Sample Size	265	214	258	217	257	207	189	171
Panel B. Second Order Polynomial								
RD Estimate	-11.946	-12.358**	8.042	9.510	-0.298	-0.431**	-0.391**	-0.433***
	(7.892)	(5.591)	(7.392)	(6.568)	(0.321)	(0.208)	(0.181)	(0.139)
Bandwidth selection	Optimal	Optimal	Optimal	Optimal	Optimal	Optimal	Optimal	Optimal
Bandwidth	12.89	14.77	13.97	10.76	11.75	14.64	12.29	13.82
Controls		X		X		X		X
Conventional p-value	0.08	0.02	0.22	0.12	0.29	0.03	0.02	0.00
Bias-Robust p-value	0.13	0.03	0.28	0.15	0.35	0.04	0.03	0.00
Sample Size	298	328	316	260	276	327	290	314

Notes: This table presents the point estimates of equation (5) using different bandwidths and polynomials on mayor's governance-related actions. The specifications in even columns include the controls displayed in Figure A8 as well as baseline version of the outcome variable. Standard errors in parenthesis are the ones estimated by the robust bias-corrected method proposed by Calonico et al. (2014); conventional and bias-robust p-values are shown in the lower part of the figure; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A28: Local Public Revenue

	(1) Total Revenue (per cap.)	(2) Taxes (per cap.)	(3) Transfers (per cap.)	(4) Other Sources (per cap.)
Panel A. Pooled Treatment Effects				
[T] Any treatment	0.047	0.016	0.003	0.054
	(0.171)	(0.040)	(0.034)	(0.128)
	[0.757]	[0.807]	[0.927]	[0.720]
Panel B. Letter vs. No Letter				
[NL] No Letter - Any Ad	0.111	0.078	-0.032	0.125
	(0.223)	(0.101)	(0.040)	(0.152)
	[0.653]	[0.570]	[0.400]	[0.550]
[L] Letter - Any Ad	0.014	-0.016	0.021	0.017
	(0.180)	(0.035)	(0.038)	(0.141)
	[0.933]	[0.700]	[0.587]	[0.913]
Test $NL = L$, p-value	0.63	0.39	0.18	0.46
Control Mean	7.597	1.315	4.042	2.255
Sample Size	685	685	679	679

Notes: This table presents the effect of the intervention on total public revenue in 2020-2023, as well specifically from national transfers, local taxes or other sources. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). A baseline version of each variable measured in 2018-2019 is included in the set of possible controls. Robust standard errors are shown in parentheses and randomization inference p-values are shown in square brackets; *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A29: Effects on Education

	(1) Facilities (Log)	(2) Teachers (Log)	(3) Students (Log)	(4) Admin. Exp. (Log)	(5) Facilities per 10k citizens	(6) Teachers per 10k citizens	(7) Students per 10k citizens	(8) Admin. Exp. per 10k citizens
Panel A. Pooled Treatment Effects								
[T] Any treatment	-0.007	-0.003	-0.003	-0.015	-0.009	-0.089	-0.632	-0.030
	(0.005)	(0.006)	(0.008)	(0.017)	(0.013)	(0.069)	(1.143)	(0.030)
	[0.207]	[0.667]	[0.650]	[0.340]	[0.497]	[0.213]	[0.587]	[0.310]
Panel B. Letter vs. No Letter								
[NL] No Letter - Any Ad	-0.010	-0.001	0.002	-0.018	-0.006	-0.072	-0.156	-0.033
	(0.008)	(0.007)	(0.010)	(0.019)	(0.016)	(0.071)	(1.472)	(0.037)
	[0.150]	[0.900]	[0.900]	[0.373]	[0.700]	[0.373]	[0.913]	[0.373]
[L] Letter - Any Ad	-0.005	-0.004	-0.006	-0.014	-0.011	-0.098	-0.880	-0.028
	(0.006)	(0.007)	(0.008)	(0.018)	(0.014)	(0.080)	(1.241)	(0.032)
	[0.340]	[0.570]	[0.467]	[0.467]	[0.417]	[0.223]	[0.507]	[0.383]
Test $NL = L$, p-value	0.50	0.68	0.45	0.82	0.75	0.72	0.61	0.87
Control Mean	3.597	5.270	7.998	3.577	2.123	9.706	152.871	1.812
Sample Size	698	698	698	698	698	698	698	698

Notes: This table presents the effects of the intervention on public education inputs measured in 2022-2023. All specifications include the covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). A baseline version of each variable measured in 2019 is included in the set of possible controls. Robust standard errors are reported in parentheses, and randomization inference p-values are reported in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A30: Effects on Healthcare

	(1) Facilities per 10k citizens	(2) Physicians per 10k citizens	(3) Nurses and assistants per 10k citizens	(4) Stretchers per 10k citizens	(5) Beds per 10k citizens	(6) Ambulances per 10k citizens	(7) Offices per 10k citizens	(8) Rooms per 10k citizens
Panel A. Pooled Treatment Effects								
[T] Any treatment	0.008	0.222	-1.563	-0.183	-0.099	0.050	-0.048	0.090*
	(0.028)	(0.546)	(1.911)	(0.201)	(0.184)	(0.048)	(0.263)	(0.051)
	[0.810]	[0.657]	[0.387]	[0.393]	[0.603]	[0.240]	[0.813]	[0.090]
Panel B. Letter vs. No Letter								
[NL] No Letter - Any Ad	0.000	0.896	1.508	0.019	-0.158	0.077	0.024	0.056
	(0.032)	(0.711)	(2.621)	(0.262)	(0.243)	(0.063)	(0.346)	(0.069)
	[0.990]	[0.253]	[0.547]	[0.953]	[0.480]	[0.273]	[0.923]	[0.430]
[L] Letter - Any Ad	0.011	-0.123	-3.148	-0.288	-0.068	0.037	-0.085	0.108*
	(0.033)	(0.546)	(1.923)	(0.216)	(0.195)	(0.051)	(0.287)	(0.057)
	[0.787]	[0.840]	[0.067]	[0.150]	[0.767]	[0.453]	[0.773]	[0.060]
Test $NL = L$, p-value	0.76	0.07	0.04	0.21	0.69	0.49	0.75	0.45
Control Mean	1.313	7.584	41.019	3.847	5.792	1.438	5.771	1.148
Sample Size	698	698	698	698	698	698	698	698

Notes: This table presents the effects of the intervention on health provision outcomes measured in 2023. All specifications include covariates selected using the method described in Chernozhukov et al. (2015) and Belloni et al. (2014). A baseline version of each variable measured in 2019 is included in the set of possible controls. Robust standard errors are reported in parentheses, and randomization inference p-values are reported in square brackets. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.