Information, oversight, and compliance: A field experiment on horizontal accountability in Brazil*

Guillermo Toral†

June 26, 2019

Latest version here

Abstract

Does the provision of information about local bureaucracies to the politicians who oversee them decrease irregularities and improve bureaucratic effectiveness? Information interventions are appealing because of their solid microeconomic foundations and their relatively low costs. However, recent experimental studies of information campaigns aimed at fostering vertical accountability (between voters and politicians) have found mixed results. Providing information to politicians directly could be more powerful, given politicians' direct responsibility for allocating and managing resources. Information may be particularly effective when provided by auditing institutions, given politicians' susceptibility to sanctions by these horizontal accountability actors. I partnered with the audit court of the Brazilian state of Rio Grande do Norte to experimentally study the effects of informing local politicians (both in government and in the opposition) about irregularities and performance in the bureaucracies they oversee. Outcomes are measured using administrative payroll data, a face-to-face survey of bureaucrats, and an online survey of politicians. Preliminary results suggest the treatment reduced the share of workers hired under temporary contracts, increased knowledge about rules among politicians, and changed politicians' sense of accountability pressure from the state audit court.

*I am indebted to Ben Ross Schneider, Daniel Hidalgo, and Lily Tsai for their feedback and advice, and to the State Audit Court of Rio Grande do Norte (TCE-RN) – in particular Gilberto Jales, Anne Carvalho, Victor Fernandes, Marcelo Araújo and Ilueny Santos – for their commitment to the project. I also thank conference participants at LASA and MPSA and seminar participants at TCE-RN and UFRN for helpful comments. Special thanks to 23 Brazilians listed in Appendix E who provided excellent research assistance for fielding the survey of bureaucrats. The project received generous financial support from MIT GOV/LAB and the Lemann Foundation. It was reviewed and approved by the Committee on the Use of Human Subjects at MIT (protocol 1810539206) and pre-registered with EGAP before data collection activities began (pre-registration 20181112AA). Any errors are my own.

†PhD candidate in Political Science, Massachusetts Institute of Technology: gtoral@mit.edu.

1 Introduction

Governments around the world have the potential to foster human development by facilitating access to universalistic, well-functioning education and healthcare services, supporting economic development, and ensuring public safety and stability. Despite significant progress in schooling, healthcare, and poverty indicators, important challenges remain across the developing world. Leakages in the management of public funds, limits to political competition, and clientelistic exchanges have consistently been associated to development failures. In this context, both researchers and policymakers are increasingly turning to governance interventions to understand and foster economic and human development. With strong foundations on microeconomic theory, and fostered by advancements in information technology, transparency initiatives have been implemented around the world to foster accountability between government and citizens. A common approach has been to provide information to citizens in order to decrease information asymmetries, improve political selection, and/or foster coordination. Nonetheless, recent experimental studies of this kind of interventions have found mixed results of citizen information campaigns on electoral accountability, with effects sometimes going in unexpected directions, or depending on mediators like coordination, expectations, and priors.[1]

In this context of increasing awareness about the challenges for vertical accountability, researchers and donors are exploring venues of horizontal accountability, i.e. accountability between governments and oversight bodies such as auditing institutions, prosecutors, and courts. Some experiments have shown that auditing institutions can decrease rent extraction of local governments.[2]

[1] Boas et al. (2018) for example find in an experiment in Brazil that voters do punish local incumbents whose accounts had been rejected by the audit court in a vignette experiment, but not when it comes to real voting behavior. Adida et al. (2016) find that informing voters in Benin about the performance of their legislators led them to *punish* good performers, rather than rewarding them, unless the information was disseminated in a way that increased its salience and facilitated voter coordination. Relatedly, Adida et al. (2017) find that ethnicity moderates the effect of information on electoral accountability: voters reward good performers only if they are co-ethnics, and punish bad performers only if they are non-co-ethnics. Arias et al. (2017) find that randomized revelations of malfeasance by mayors in Mexico *increased* support for the incumbent party, which the authors attribute to pre-treatment low prior beliefs among many voters. In a similar vein, Gottlieb (2016) finds in an experiment in Mali that raising voters' expectations of government makes them more likely to hold politicians accountable. Finally, Buntaine et al. (2018) find that Ugandan voters who receive information about local councilors irregularities punish bad performers only when they are running for lower-level positions.

[2] For example, Avis et al. (2017) exploit Brazil's randomized audits and find that these audits reduce corruption in subsequent years. Zamboni and Litschig (2018) present results of a field experiment in that same setting and show that increasing the risk of audit reduced corruption in procurement. The

Powerful audits are however costly and hard to scale up. Information interventions, on the other hand, are significantly cheaper but also have the potential to increase perceptions of accountability pressure, decrease information asymmetries, and foster accountability behaviors among politicians. Unlike citizens, elites may be particularly responsive to information given their positions of direct responsibility, in at least two ways. First, in contexts with adequate institutions of horizontal accountability,[3] they can be punished for malfeasance on the grounds of electoral, civil and even penal law (e.g. loss of their position, payment of fines, and imprisonment, respectively). Second, political elites have the tools to directly affect governance outcomes, be it through resource allocation or management decisions. An emerging literature has started to use experiments to examine the effects of providing information to politicians on governance. For example, Raffler (2018) provided Ugandan politicians with information and training on their oversight responsibilities, and found positive effects on oversight among local politicians in subcounties that were not aligned with the central government. In Peru, Lagunes (2017) examined the effect of sending letters to municipal governments announcing monitoring of civil works projects by state and non-state actors, and found significant decreases in project costs.

This project contributes to the emerging experimental literature on the dynamics of within-government accountability. In partnership with the audit court of the Brazilian state of Rio Grande do Norte, I randomly assigned half of 164 municipalities to a treatment group where local politicians (both in government and in the opposition) received a scorecard highlighting bureaucratic irregularities, bureaucratic performance, and politicians' responsibilities in bureaucratic oversight. I hypothesized that the treatment would increase accountability inputs (knowledge, self-efficacy, and perceptions of accountability pressure among politicians), accountability outputs (accountability pressures between the executive and the legislative and between the executive and the bureaucracy, as well as political constraints on patronage), and accountability outcomes (compliance with hiring rules and bureaucratic performance). Primary outcomes are measured through detailed administrative data (monthly payroll data), and mechanisms are measured with original surveys of bureaucrats (face-to-face) and politicians (online). The experimental design was based on continuous dialogue with the field partner, Rio Grande do Norte's state audit court (*Tribunal de Contas do Estado do Rio*

effectiveness of audits however may depend on institutional arrangements like their strength or timing. Bobonis et al. (2016) find that in Puerto Rico audits decrease corruption only when they are scheduled before elections, and De la O and García (2015) find that federal audits (but not state audits) lead to changes in subsequent municipal spending.

[3]As notably argued by O'Donnell (1998), the effectiveness of horizontal accountability institutions requires them to be both authorized and willing to act, as well as to operate within a network of of accountability agencies.

Grande do Norte, TCE-RN). TCE-RN is a highly capable institution with an interest in developing evidence-based strategies for guiding their work in overseeing and auditing government institutions in the state. The design also built on over 120 in-depth interviews I have done across 6 Brazilian states – including Rio Grande do Norte – with municipal bureaucrats, municipal politicians, and state prosecutors in charge of overseeing municipal governments. Before any data collection began, a detailed pre-analysis plan was registered with EGAP – any deviations from this plan are noted in the paper and detailed in Appendix J.

Preliminary results show that the treatment decreased the share of municipal employees who are hired under temporary contracts in average by about 0.16 standard deviations ($p < 0.05$). Survey data suggest this effect was driven by an increase in politicians' knowledge about rules (which increased, in average, by about 0.12 standard deviations in treated municipalities, $p < 0.1$). Perceptions of accountability pressure from the state audit court appear to be *lower*, rather than higher, among politicians in treated municipalities (the difference is of about 0.16 standard deviations, $p < 0.05$). Some evidence suggests that this is a manifestation of politicians feeling increased oversight and surveillance, since similar treatment effects can be detected in perceptions of accountability pressure of the state prosecutor's office, which was not involved in any way in the experiment. Additionally, some evidence suggests the treatment induced changes in the relationship between the executive and the legislative. In treated municipalities, mayors and secretaries report significantly less meetings with city councilors (-0.22 standard deviations, $p < 0.01$), and both them and bureaucrats have worse perceptions of the role of city councilors in service delivery. These changes could help explain the changes in the hiring of municipal employees, given the role that city councilors have traditionally played in fostering patronage in small and medium municipalities of Brazil like the ones I study.

The experiment has some important implications for policy making, in Brazil as well as other settings with strong horizontal accountability institutions. Auditors, prosecutors and other anti-corruption agents around the world have increasing access to detailed, administrative data about what governments do, including data on hiring and procurement decisions. These institutions however often lack evidence-based guidance on how to leverage this data to fulfill their mission as accountability actors.Results suggest that even low-cost interventions can improve information levels among local elites and increase compliance with rules. More research however is needed to fully understand how messages from auditing institutions can affect dynamics of within-government accountability at the local level.

The rest of the paper is organized as follows. Section 2 discusses the empirical setting, with particular attention to municipal governance institutions in Brazil and the specifics of Rio Grande do Norte. Section 3 discusses the experiment's research design, providing details about the treatment, the randomization procedure, and the sampling strategy. Section 4 presents the experiment's theory of change and the corresponding hypotheses. After briefly discussing the data in Section 5, I present the estimation strategy in Section 6, discussing issues of inference, multiple testing, and power. Section 7 presents preliminary results from a completed face-to-face survey of bureaucrats (n=926, over 90% of the target population), and an ongoing online survey of politicians (n=905, about 33% of the target population). Section 8 briefly concludes with a discussion of these preliminary findings and next steps.

2 Context

Brazil is a large, highly decentralized, middle-income country where local governance is particularly critical for development, given municipal governments' responsibility in the provision of public services. In fact, despite some convergence of less-developed regions (like Northeastern states), there remains wide variation in development outcomes, even among municipalities with similar levels of income. The coexistence of formally equivalent government structures across the country with wide variation in outcomes, and the availability of rich, micro-level datasets make Brazil a suitable context for studying within-government accountability.

2.1 Local governance in Brazil

Brazil's 1988 constitution established a three-level federal system and gave significant fiscal autonomy to municipal governments. Among other things, municipalities are responsible for providing primary education and healthcare, two areas in which they are obliged to spend at least 40% of their revenue, as well as primary social assistance.[4] To achieve those goals, municipalities hire large numbers of bureaucrats – in fact, they spend in average about 60% of their income in salaries. Municipalities can hire bureaucrats essentially under three types of contracts: civil service positions (tenured positions for which candidates must pass a competitive examination), temporary positions

[4]Municipal governments, however, depend heavily on inter-governmental transfers and raise only a small fraction of the revenue they spend (Arretche, 2004).

(one-year positions which are supposed to be assigned after some selection process but are often assigned ad hoc), and positions of trust (for which politicians have most discretion – these are positions that are to be used only for leadership, direction and advisory roles but which are sometimes abused for other purposes). As per the constitution, civil service hiring must be the default for any permanent staffing needs, such as teachers or doctors, but in practice this varies widely. A number of clear legal requirements are also commonly violated across the country, including the limit on personnel expenses (which are not supposed to go over 54% for the executive), the need to do some selection procedure also for temporary hires, and the ban on any person holding more than two government jobs. In general, observers have long noted the prevalence of patronage in municipal bureaucracies in Brazil, something that recent research has shown as well (Akhtari et al., 2018; Colonnelli et al., 2018; Toral, 2019b).

Elections take place every four years, with municipal elections taking place two years before/after state and federal ones. The last municipal elections took place in 2016. Municipal elections consist of simultaneous elections for a mayor (who is elected through a majoritarian system) and for a number of city councilors that depends on the municipality's population (and who are elected though a proportional, open-list system).[5] Mayors, who appoint a set of non-elected secretaries, need to build and sustain a base in the legislative chamber in order to pass legislation (including the yearly budget law) and to avoid being impeached.[6]

This institutional design, together with high levels of political competition, produces local dynamics of what has been conceptualized at the federal level as "coalitional presidentialism" (Abranches, 1988; Power, 2010). While little is known about the dynamics of legislative-executive relations at the local level, my fieldwork suggests there is wide variation in the extent to which mayors have control over (or the support of) the chamber. Still, city councilors generally have considerable leverage in the allocation of patronage positions in the municipal bureaucracy. This is sometimes discussed openly by local politicians during in-depth interviews. For example, a secretary of administration (i.e., in charge of human resources) in a municipality of Rio Grande do Norte said, when asked about the role that city councilors play in appointments:

[5]The number of city councilors to be elected in each municipality is regulated by a municipal law, subject to limits established by the federal constitution. For municipalities in Rio Grande do Norte, the *maximum* number of city councilors established by the constitution ranges from 9 to 29.

[6]Passing legislation usually requires a simple or absolute majority, while an impeachment requires a qualified majority of two thirds of the chamber. City councilors also play a key role in the review process of municipal accounts – if the state audit court rejects the accounts, these can still be approved with the vote in favor of two thirds of the chamber.

"Politics goes hand in hand with management, because the government depends on the support of the [legislative] chamber. I listen to petitions from city councilors within the financial capacity [of the government]. For example, if there is an opening for a guard, I listen to to [city councilors'] recommendations. But it is not only city councilor requests [that counts], they must also submit a CV, have experience... We do not allocate jobs slapdash."

Brazil has a rich environment of horizontal accountability institutions (O'Donnell, 1998), including state audit courts and powerful and autonomous state prosecutor offices. These institutions, which have become increasingly powerful over time (Praça and Taylor, 2014), act on strong legal bases, which allow for judges to terminate political mandates, to impose strict fines, and even to imprison malfeasant politicians. State audit courts, which are in charge of overseeing accounts of municipal and state governments, play an important role in this system of horizontal accountability. They are in charge of reviewing municipal (and state) government accounts,[7] and can conduct audits, impose sanctions (like fines), suspend bidding processes, and report any malfeasance they detect to state prosecutors, who could take politicians to court. Researchers have documented that state audit courts in Brazil have varying degrees of dependency on the executive (Melo et al., 2009). While the appointment system of state audit court councilors (Hidalgo et al., 2016) and more informal links between councilors and politicians (Sakai and Paiva, 2016) limit the independence of these institutions from state governments, the fact that they are increasingly large and professionalized bureaucracies, and that they combine their legal competencies with softer actions like media outreach, trainings, and data collection and dissemination makes them generally powerful actors particularly with respect to governments of small and medium-sized municipalities.

2.2 Rio Grande do Norte

The intervention took place in Rio Grande do Norte (RN), a state at the heart of the Northeastern region. The Northeast is Brazil's poorest region, and has historically been characterized by inferior development outcomes, corruption, and clientelistic politics, particularly in the allocation of public jobs (Leal, 1948). I chose RN largely due to the state audit court's willingness to partner for an

[7]If the municipal accounts are rejected by the state audit court, and the rejection is not overcome by a qualified majority of the municipal chamber, the mayor becomes ineligible for running in any election for the following eight years.

academic study, and to its access to monthly payroll data of all municipal governments in the state. RN is also convenient in that its relatively small size limits the costs of implementing an in-person survey. Rio Grande do Norte has 167 municipalities, of which only 3 have more than 200,000 inhabitants – the rest are quite small, with between 1,600 and 81,000 inhabitants, and a median population below 10,000. In average these municipalities have one third of their population living in rural areas, which generally depend either on subsistence agriculture or agricultural work for large companies.

Municipalities in RN exhibit wide variation across a range of socioeconomic and political variables, as illustrated in Appendix A. Like much of the Northeast, municipalities in RN generally have less people and are poorer when compared to municipalities in the whole of Brazil. Partly as a result of these two facts, municipalities in RN tend to have a larger share of the population hired as municipal employees. These municipalities generally lag behind in human development outcomes such as infant mortality rates, student learning, and student passing rates. Despite persistent challenges of clientelism, elections are generally competitive.

Like all states in Brazil, Rio Grande do Norte has its own state audit court (*Tribunal de Contas do Estado de Rio Grande do Norte*, TCE-RN), which in 2018 had 436 staff (including the constitutionally mandated 7 councilors) and got in the 2017 budget law 0.7% of the state's budget (over 90 million Brazilian reais, or about USD28 million). TCE-RN's self-declared mission is"to exercise external control, guiding and supervising the use of public resources, to the benefit of society." To achieve that mission, it aims at (i) "curbing the occurrence of fraud and the diversion of public resources;" (ii) "contributing to the improvement of performance and transparency in public management;" and (iii) "increasing the effectiveness of external control actions" (TCE-RN, 2018). This experiment aims at helping TCE-RN achieve all of these targets.

3 Research design

3.1 Treatment

The intervention consisted of a municipal bureaucracy scorecard sent to key politicians in the executive branch (the mayor and some of the secretaries appointed by them, namely those of finance,

management,[8], education, healthcare, and social assistance), and all politicians in the legislative branch (city councilors, including those in the mayor's coalition and those in the opposition). The scorecard was printed as an A3-sized poster in color in high-quality paper, with the state audit court's logo and an accompanying letter signed by its president.[9] The poster had the following municipality-specific data, presented in easy-to-understand graphs.[10]

- Descriptive statistics of the local bureaucracy. These include the number of municipal employees and the amounts spent in their salaries, disaggregating them by contract type (tenured, temporary, positions of trust, and other). These graphs leverage monthly payroll data submitted to TCE-RN by municipal governments every month. The scorecard also mentioned the share of the municipality's revenue that was spent in salaries in 2016, leveraging TCE-RN data.

- Performance statistics of the municipal government. This include key statistics about the municipal systems of basic healthcare and basic education, leveraging administrative data from the federal government. It also included municipality-level performance statistics from TCE-RN's Municipal Management Effectiveness Index (IEGM[11]), which assesses the municipal government's effectiveness overall and in 7 policy areas: education, healthcare, management, finances, environment, urban planning, and transparency.

The scorecard also included reminders about the most important legal constraints in hiring (e.g. prevalence of civil service hiring for permanent staffing needs, the general prohibition of anybody holding more than one government job, the limited circumstances in which workers can be hired under temporary contracts or positions of trust, and the limit on the share of municipal revenue that can be spent on personnel). It also emphasized politicians' responsibility in ensuring compliance with the law and in making sure that the municipality's investment in salaries leads to better public services for all. Finally, the accompanying letter made it clear what other actors were receiving the information in order to foster common knowledge.

Scorecards were sent twice: hard copies by registered mail in late November, and electronic copies by e-mail in late January. The January version of the scorecard was updated to reflect the

[8]Municipal secretaries of management (*secretários de administração*) are in charge of managing the municipality's human resources.

[9]The scorecard and the survey instruments were piloted in municipalities in Minas Gerais and Ceará.

[10]A sample of the scorecard can be seen in Appendix B.

[11]https://iegm.tce.rn.gov.br/

newest IEGM data. The timing was decided to ensure that politicians receive the information in a timely manner, close to but ahead of the moment when many hiring decisions are made.[12]

3.2 Randomization

Of the 167 municipalities in Rio Grande do Norte, 4 had to be excluded from the experiment because they do not submit payroll data to the state audit court. Since a pairwise randomization strategy requires an even number of units, I further exclude the state capital – Natal – which is in many ways different from all the other municipalities in the state (four times larger than the second largest, a much richer information environment, and way more complex governance structures than in typical municipalities of the interior).

The resulting 162 municipalities that compose the experimental sample were randomly assigned to a treatment condition (where politicians received the scorecard) or a control condition (where politicians received nothing).[13] In order to increase statistical power, which was a major concern given the small sample size, randomization followed a paired matched design. Pairs were created using the machine learning algorithm of Barrios (2014), which is optimal in minimizing the mean squared error of the difference in means between treated and control units.[14] Observable pre-treatment characteristics of municipalities in treatment and control groups are well balanced.[15]

[12]Municipal governments generally end temporary contracts in December, and start them again in the first few months of the following year (Toral, 2019b).

[13]Note that politicians in all municipalities in the state regularly receive information and requests from TCE-RN, and thus an additional placebo was not necessary.

[14]This algorithm follows four simple steps. First, I used ten-fold cross-validated LASSO (the least absolute shrinkage and selection operator) to generate a prediction function where the dependent variable is a baseline measure of my primary outcome (the share of municipal employees who are tenured). Second, I used predicted values of the outcome from the sparse conditional expectation function generated with LASSO in order to rank municipalities. Next, I created pairs following that rank, such that the first two municipalities form a pair, the next two form another pair, etc. Finally, I randomly assigned treatment within pairs using a random number generator in R.

[15]Examining differences between treatment and control municipalities along 45 covariates (of which only 7 are in the LASSO-generated sparse conditional expectation function used to generate pairs), using both difference-in-means and Kolgomorov-Smirnov tests, the lowest p-value is 0.07. A balance table is included in Appendix G.

3.3 Sampling and survey implementation

Rio Grande do Norte has 167 municipalities. The state capital Natal and 4 municipalities that reported no payroll data at baseline were excluded from the experimental sample. The remaining 162 municipalities were assigned half to treatment and half to control. All municipalities -including Natal- are being targeted in the online survey of politicians. On the other hand, the largest 17 municipalities in the state (which had as of the 2010 census more than 30,000 inhabitants) were excluded from the field survey of bureaucrats, for two main reasons. First, because of their size including them in the sampling frame would heavily increase the cost of the survey. Second, including them in the sampling frame would expose enumerators to the serious security challenges typical of large urban areas of the Northeast.

All politicians listed in Section 3.1, in both treatment and control municipalities, were sent the politician survey through TCE-RN's online accountability system, as a strategy to improve response rates. As for the field survey of bureaucrats, I targeted all "street-level managers" of municipal pre-schools and basic education schools, basic health clinics,[16] and basic social assistance centers[17] in urban areas[18] of the 150 largest cities in the state. Of a target population of 1,027 street-level managers, 926 were surveyed.[19]

The experiment experienced some implementation delays, particularly with regards to the sending of scorecards and the launch of the online survey of politicians, which only started in mid February.[20] The scorecards were sent in hard copies in late November, and then again by e-mail (with updated IEGM data) in late January. The face-to-face survey of bureaucrats was completed

[16] *Unidades básicas de saúde*

[17] *Centros de Referência da Assistência Social*

[18] Schools, clinics and social assistance centers in rural areas were excluded from the sampling frame, for three main reasons. First, rural schools and clinics are often staffed for a limited number of days and hours per week. Second, the directors of rural schools very often work at the municipality's urban center, and tend to direct several schools at once. Third, rural areas in the Northeast are logistically hard to reach – they are often accessible only through dirt roads with limited or no GPS service, unmapped on GPS services like Waze or Google Maps. Therefore, including rural areas in the sampling frame would heavily increase the time and budget required for the survey, and may raise security issues for enumerators. While there are many schools and clinics in rural areas, most of the population lives in urban areas and is thus served by urban schools and clinics. For example, while over 55% of the 2,415 municipal schools in Rio Grande do Norte are in rural areas, they concentrate less than 27% of municipal student enrollments in basic education.

[19] See Figure H.5 in the Appendix for details on survey completes by municipality.

[20] For details on the planned and the actual implementation calendar, see Figure H.6 in Appendix H.

in a timely manner. The online survey of politicians was however launched later than expected, and is experiencing slower responses than we had anticipated. Since it was launched in mid February, 905 have responded (out of a target population of about 2,800 – all mayors, city councilors, and secretaries of the key five areas indicated in Section 3.1). While attrition is not correlated with treatment assignment, low response rates make outcome data noisier and limit the number of units that can be used in municipality-level regressions. Online and telephone reminders have been sent.[21] As for administrative, payroll data, TCE-RN has now complete payroll data for the month of February for 164 of the state's 167 municipalities for the month of February.[22]

4 Hypotheses

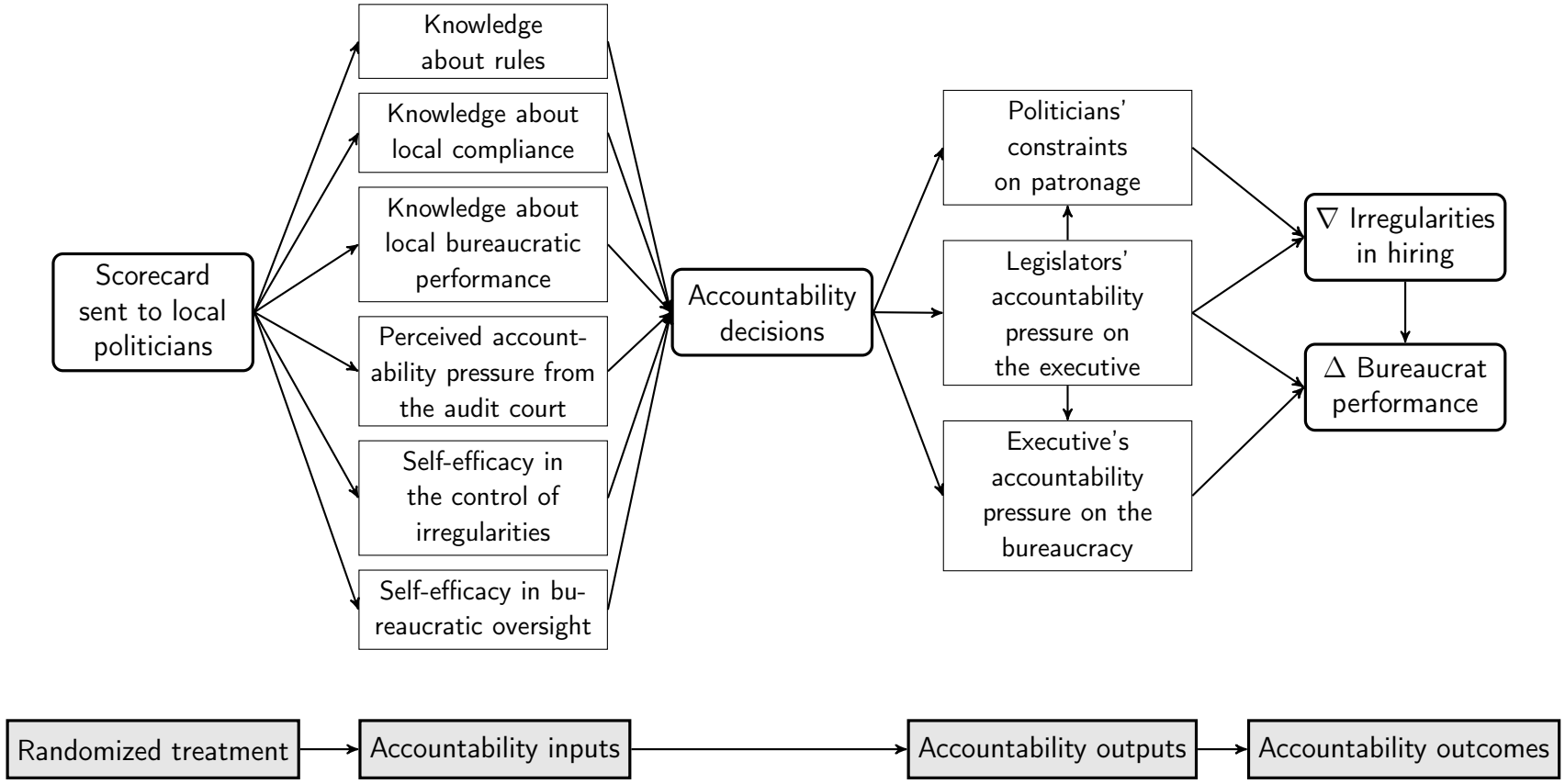
Figure 1 visually represents the experiment's theory of change and synthesizes the hypotheses. At the highest level, I hypothesized that the randomized scorecard would improve what I call accountability inputs among politicians (i.e. knowledge, perceived accountability pressure, and self-efficacy); that these in turn would lead to changes in what I call accountability outputs (i.e. politicians' behaviors vis-a-vis each other and the bureaucracy); and finally that these would improve accountability outcomes (i.e. development goods). Figure 1 thus represents the whole hypothesized chain linking treatment to outcomes.

Separating both in theory and measurement these different accountability inputs and outputs is useful in at least two ways. First, measuring treatment effects on these intermediate variables is helpful to empirically address the mechanisms behind potential treatment effects on accountability outcomes. Second, measuring treatment effects on different accountability inputs (knowledge, perceptions of accountability pressure, and self-efficacy) can give us a sense of the constraints in within-government accountability that can be eased through this type of intervention. In both cases, measuring the effect of treatment in these accountability inputs and outputs can provide insights into how the experiment worked (or why it did not), as well as guidance for future policy and research.

[21] As an additional incentive, we will be sending municipality-specific reports in a few months, such that politicians can compare how they fare (e.g. in terms of knowledge about rules or compliance, and perceptions of local government stakeholders) vis-a-vis other municipalities in the state.

[22] I pre-specified outcomes would be analyzed using March payroll data, since many temporary positions in the education sector are only filled in then. I will therefore re-analyze the data once the March payroll data is available.

Figure 1: The experiment's theory of change



Additionally, I pre-specified some hypotheses about heterogeneous treatment effects, namely that increases of accountability outputs and outcomes would be higher in municipalities where a minority of city councilors support the mayor, where the mayor is in their first term, or where the share of municipal employees who are tenured is below the median. At the individual level, I hypothesized increases in accountability inputs would be higher among city councilors in general or among opposition city councilors. Finally, I hypothesized increases in accountability pressure on bureaucrats would be larger among politically connected bureaucrats.

More details on hypotheses and the variables in survey and administrative data that they are linked to are included in Appendix C. In general, hypotheses about accountability inputs and outputs are measured using the surveys of bureaucrats and politicians. Hypotheses about irregularities in hiring are tested using administrative payroll data submitted by municipal governments to TCE-RN. Hypotheses about bureaucratic performance will be tested using administrative performance data, once they are released by the federal government in 2020.

5 Data

The experiment examines several kinds of outcomes:the size, distribution, and compliance of local bureaucracies; informational and attitudinal outcomes; and the performance of local bureaucracies. Statistics about the local bureaucracy (numbers and types of contracts, expenses on personnel) come from monthly, administrative payroll data on all municipal contracts submitted every month by municipal governments to TCE-RN.[23] Statistics about the performance of local bureaucracies will be obtained next year from administrative records maintained by the federal government.

Data on intermediate outcomes comes from two surveys. With the help of 23 research assistants, I implemented a face-to-face survey of bureaucrats in late 2018. We interviewed over 90% of the street-level managers in urban areas of the 150 largest municipalities in the state. The median number of bureaucrats surveyed per municipality is 4.[24] The online survey of politicians is currently being implemented, together with TCE-RN, targeted at mayors, secretaries and city councilors in all municipalities in the state. So far, 905 politicians have responded, or about 33% of a target

[23]TCE-RN's Resolution 030/2012 mandates that all municipal executive and legislative powers submit their payroll to the court every month, and allows for the imposition of fines to governments that do not comply.

[24]More details about the survey of bureaucrats are included in Toral (2019a).

population of about 2,800 (all mayors and city councilors in the state, plus secretaries of 5 key areas). By type of actor, response rates are over 38% among mayors, over 37% among presidents of legislative chambers, and over 64% among secretaries. The median number of politicians surveyed per municipality is 5. Attrition is not correlated with treatment assignment, as shown in Table 1 below. Nonetheless, low response rates make outcome data noisier and limit the number of units that can be used in municipality-level regressions. A number of actions are therefore being taken to boost response rates, as detailed in Section 3.3. Links to the survey questionnaires are included in Appendix D.

Table 1: Survey response rates are not correlated with treatment

	Dependent variable: Number of respondents			
	Politicians survey		Bureaucrats survey	
	(1)	(2)	(3)	(4)
treatment	0.975	0.975	0.506	0.506
	(0.721)	(0.738)	(0.518)	(0.543)
Randomization pair fixed effects	✓		✓	
Observations	162	162	162	162
R^2	0.527	0.011	0.547	0.005

*p<0.1; **p<0.05; ***p<0.01

Pre-treatment covariates used in the LASSO regression for pairing municipalities came from multiple sources, including TCE-RN, the Supreme Electoral Court, the official statistics institute IBGE, UNDP, and the Ministries of Finance, Education, Healthcare, and Social Development.

6 Estimation and inference

The experiment collects both municipality-level and individual-level data, from administrative and survey sources. Some of the experiment's outcome variables (those corresponding to the accountability outcomes as defined in Section 4) are inherently municipal. Others however are individual-level variables that can either be aggregated up to be analyzed at the level of the municipality, or analyzed at the individual level, which has advantages of increasing power and allowing for the

examination of individual sources of heterogeneity in treatment effects.

My estimand of interest is the average treatment effect, or difference between expected potential outcomes of municipalities under treatment and under control: $\tau = \mathbb{E}[Y_{m,1}] - \mathbb{E}[Y_{m,0}]$. Since treatment assignment is randomized and is thus independent of potential outcomes, the average within-pair difference in observed outcomes is an unbiased estimator of the average treatment effect. Paired matched designs are in essence a block designed where blocks are composed of two units (one of which is randomly assigned to treatment). I therefore estimate the average treatment effect regressing the outcome data at the municipality level on a treatment indicator and a set of pair dummies:

$$Y_{mp} = \sum_{j=1}^{J} \alpha_j \mathbf{I}[j = p] + \beta D_{mp} + \theta(\tilde{Y}_{mp} - \bar{\tilde{Y}}_{mp}) + \gamma(\tilde{Y}_{mp} - \bar{\tilde{Y}}_{mp})D_{mp} + \varepsilon_{mp} \tag{1}$$

where Y_{mp} is the outcome of interest for municipality m in randomization pair p, $\sum_{j=1}^{J} \alpha_j \mathbf{I}[j = p]$ is the set of J pair dummies (one of which acts as the intercept), and D_{mp} is the treatment indicator $\in \{0, 1\}$ for municipality m in pair p. $\tilde{Y}_{mp}$ is a baseline measure of the outcome, which I include demeaned and interacted with treatment, for precision, in regressions with administrative data for which I have baseline measures.[25] For inference about the null hypothesis of no average treatment effect, I use HC2 heteroskedasticity consistent standard errors in municipality-level regressions with payroll data. In regressions with individual-level survey data I use municipality-clusteted standard errors since the design imposes clustering in treatment assignment (**?**).

I also test the sharp null of no effect for all pairs ($Y_{j,1} = Y_{j,0} \forall j$) using randomization inference. To do so, I will derive the full schedule of potential randomizations (under the pairwise randomized design, and without changing the pairs that will be determined before the actual randomization, as discussed in Section 3.2) and simulate the average treatment effect under all of them to derive the sharp null distribution of $\hat{\tau}$.

I address the multiple testing issue (due to the fact that for each hypothesis listed in Section 4 I use a number of variables) with two strategies. First, for survey data, which I use to test hypotheses 1 through 9, I generate one index for each hypothesis by applying principal component analysis to

[25] By controlling for the baseline outcome I seek to increase power, which could be lower using a first-differenced design (McKenzie, 2012). Demeaning covariates and interacting them with treatment ensures unbiasedness of $\hat{\beta}$ for estimating the average treatment effect (Imbens and Rubin, 2015).

each set of survey items corresponding to one hypothesis (e.g. knowledge about rules). This serves two purposes. First, it reduces the noise-to-signal ratio, thus increasing power. Second, it aggregates multiple variables into a single one, eliminating the need for multiple testing corrections. Generating an index for each of these hypotheses is also sensible since I do not have a priori reasons to expect some survey items to respond differentially to treatment than others.Second, for administrative outcome data, which I use to test hypotheses 10 and 11, I do not aggregate different variables into an index – instead, once I have final outcome data I will use the Westfall-Young bootstrapping procedure for controlling the family-wise Type I error rate (Westfall and Young, 1993).

7 Preliminary results

This section presents results using the data available so far, namely payroll data corresponding to February 2019, and survey data for the completed survey of bureaucrats and the ongoing survey of politicians. Once I have payroll data for March and final survey data I will re-do analyses, incorporating p-values from randomization inference and multiple testing corrections.

Table 2 presents results for the effect of treatment on hiring irregularities, using February payroll data.[26] Using this data, the scorecard had no significant average treatment effect on the study's pre-registered main outcome, the share of municipal employees who are hired without tenure. The difference however goes in the expected direction: treated municipalities appear to have a smaller fraction of their bureaucrats under non-tenure contracts. I pre-specified three outcomes as components of the experiment's main outcome: the share of workers with temporary contracts, the share of workers under positions of trust, and the logged total number of workers. For the first of these three, we observe a statistically significant average treatment effect. Municipalities that received the scorecard have, in average, 2.7 percentage points less of their working force employed with temporary contracts, which is equivalent to about 0.17 standard deviations ($p < 0.05$). There is no significant effect on the total amount spent in salaries. The treatment appears to have significantly reduced the number of workers with more than two jobs in the municipality (0.96 standard deviations, $p < 0.01$). However a descriptive examination of the data suggests there may be data entry errors in that field so that result is likely to change with final data.

[26] Treatment effects on the other accountability outcome, namely bureaucratic performance, can only be measured next year once the federal data are published.

Table 2: Treatment effects on hiring irregularities, measured with February 2019 payroll data

	Dependent variable:					
	Not tenured workers (%)	Temporary workers (%)	Trust workers (%)	Total workers (log)	Total salaries (log)	Workers with > 2 job
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\beta}$: treatment	−0.020	−0.027**	0.009	−9.921	0.157	−3.132***
	(0.017)	(0.013)	(0.010)	(16.468)	(0.039)	(0.439)
	0.236	*0.037*	*0.358*	*0.646*	*0.247*	*0.076*
N	162	162	162	162	161	162
R^2	0.844	0.869	0.829	0.989	0.953	0.505

All regressions control for randomization pair fixed effects, a baseline measure of the outcome and its interaction with treatment, as per Equation 1.
HC2 standard errors in brackets below coefficients. *p<0.1; **p<0.05; ***p<0.01
Randomization inference p-values for the sharp null hypothesis in italics below standard errors

To explore mechanisms, I leverage the completed face to face survey of bureaucrats and the ongoing online survey of politicians. Table 3 presents the average effect of treatment on politicians' knowledge about rules. These results suggest the scorecard improved politicians' knowledge about the rules governing bureaucratic hires. The principal-component index is about 0.14 standard deviations higher among politicians in treated municipalities ($p < 0.05$). This effect appears driven by increased knowledge of the scenarios in which temporary hiring can be used (0.13 standard deviations, $p < 0.05$), and of the personnel spending limit (0.17 standard deviations, $p < 0.05$). I also report average treatment effects on the average of the items for each hypotheses, as an alternative to the principal component analysis approach. While this strategy for aggregating survey answers was not pre-specified, it can be a useful robustness check that is more readily transparent than the PCA index. In this case, the average score is 0.14 standard deviations higher among politicians of treated municipalities ($p < 0.05$).

Preliminary results show no statistically significant results on politicians' knowledge about the municipality's compliance with hiring rules (hypothesis 2), or about the performance of the municipal bureaucracy (hypothesis 3), as shown in Appendix I.[27] The survey of politicians does show significant results on perceptions of accountability pressure from the state audit court, but in the direction opposite to that hypothesized. As shown in Table 4, politicians in treated municipalities

[27]Note however that about half of survey participants chose not to answer these questions, so null effects may well be driven by insufficient power. Effects go in the hypothesized direction.

Table 3: Treatment effect on politicians' knowledge about rules around hiring (hypothesis 1)

	Dependent variable:				
	Knowledge about rules on:				
	temporary hires	positions of trust hires	personnel spending	PCA index	Average score
	(1)	(2)	(3)	(4)	(5)
$\hat{\beta}$: treatment	0.065**	−0.003	0.075**	0.170**	0.137**
	(0.030)	(0.033)	(0.035)	(0.086)	(0.068)
N	870	870	870	870	870
R^2	0.119	0.129	0.171	0.124	0.124

All regressions control for randomization pair fixed effects
Standard errors clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01

report significantly lower levels of agreement with the statement "TCE-RN detects irregularities in this municipality"[28] (0.28 standard deviations, $p < 0.01$). Effects on agreement with the statements "I trust TCE-RN" and "TCE-RN knows the reality of this municipality" are also negative yet statistically insignificant. Using the PCA index, the treatment depressed perceptions of accountability pressure from TCE-RN by 0.18 standard deviations ($p < 0.05$).[29]

Why would the scorecard worsen perceptions of accountability pressure from the state audit court? One option is that politicians who receive the scorecard engage in Bayesian updating: perhaps they envisioned a more capable court before receiving the poster, which somehow lowers their perceptions of the court. The most obvious scenario in which this would happen is if the data included in the scorecard were inaccurate. While the contract data that were used to create the scorecards may have measurement error, they are contract-level, administrative data reported by municipal governments to the state audit court. Moreover, qualitative interviews during the piloting of the scorecard in the state of Minas Gerais and after the experiment in Rio Grande do Norte suggest that politicians trust and value the information contained in the scorecard. An alternative hypothesis is that politicians exposed to the treatment feel intensified surveillance, and react to it by expressing lower views of the legitimacy of the court as an oversight body. Consistent with this

[28]Agreement was measured on a 4-point scale ("not at all", "a little", "quite", or "a lot"). Results are similar when recoding the variable as binary.

[29]Results are very similar using the average score.

Table 4: Treatment effect on politicians' perceptions of accountability pressure from the state audit court (hypothesis 4)

	Dependent variable: Agreement with statements about the state audit court				
	I trust TCE-RN	TCE-RN knows this municipality	TCE-RN detects irregularities in this municipality	PCA index	Average score
	(1)	(2)	(3)	(4)	(5)
$\hat{\beta}$: treatment	−0.032	−0.091	−0.232***	−0.247**	−0.118**
	(0.043)	(0.061)	(0.072)	(0.103)	(0.047)
N	771	771	771	771	771
R^2	0.137	0.129	0.157	0.145	0.147

All regressions control for randomization pair fixed effects
Standard errors clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01

latter possibility, the treatment also appears to depress perceptions of accountability pressure from the state prosecutors' office, which is a totally separate, highly autonomous institution that was not related to the treatment.[30]

Pre-registered tests of treatment effects on self-efficacy examine how respondents place themselves on questions that asks respondents to rank 7 local actors on their responsibility in the control of irregularities (hypothesis 5) and in the improvement of public services (hypothesis 6). As shown in Table A5 in Appendix I there is no evidence in favor of either hypothesis. Nonetheless, exploratory analyses looking at treatment effects on the position given to each of those actors do show some significant differences between treated and untreated politicians, which shed additional light on the effect that the treatment may have had on external accountability pressure. As shown in Table 5, politicians in treated municipalities give more responsibility in the control of irregularities to external control bodies (TCE-RN and the Prosecutors office), and less responsibility to the mayor. This evidence is consistent with the treatment increasing politicians' sense of surveillance and pushing them to express attitudes consistent with blame avoidance.

As for the effect of treatment on legislators' accountability pressure on the executive, results

[30]These results, which were not pre-registered but help to explain these counterintuitive findings, are shown in Table A4 in Appendix I.

Table 5: Treatment effect on politicians' perceptions of responsibility of different actors in the control of irregularities (not pre-registered)

	Dependent variable: Ranking in responsibility for controlling irregularities given to...						
	TCE-RN	Prosecutors	Mayor	City Councilors	Secretaries	Workers	Citizens
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	−0.209*	−0.341***	0.584***	−0.048	0.051	0.040	−0.077
	(0.120)	(0.128)	(0.103)	(0.107)	(0.106)	(0.099)	(0.132)
N	795	795	795	795	795	795	795
R^2	0.114	0.116	0.147	0.120	0.147	0.095	0.147

All regressions control for randomization pair fixed effects
Standard errors clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01

go counter to what I had hypothesized. If anything, the treatment appears to have decreased city councilor's pressure on mayors, secretaries, and bureaucrats, as shown in Table 6. Mayors and secretaries in treatment municipalities report less meetings with city councilors (0.3 standard deviations, $p < 0.05$). Mayors, secretaries and bureaucrats in treatment municipalities report worse perceptions of city councilors' information about public services and their role in their improvement.[31] One possible interpretation of these findings is that actors in the executive are responding to the increased surveillance by implicitly blaming legislators for the state of affairs.

Finally, the data show no evidence of the treatment increasing executive politicians' accountability pressure on the bureaucracy, as shown in Tables A6 and A7 in Appendix I.

8 Conclusion

Horizontal accountability actors such as auditing institutions, prosecutor offices, and courts have increasing access to administrative, micro-level data on government inputs (in areas like employment and procurement) and outputs (e.g., on public service delivery and human development outcomes). A number of these institutions are increasingly leveraging those data in attempts at overseeing public officials and reducing malfeasance. While recent research has widely explored the conditions

[31] Analyses of treatment effects on PCA and average indices for hypothesis 8 are still to be included.

Table 6: Treatment effect on legislators' accountability pressure on the executive (hypothesis 8)

	Dependent variable:								
	Meetings	Meetings	Contacts	Meetings	Meetings	Agreement with: "City councilors...			
	with	with	from	with	with	know services"		improve services"	
By:	Exec.	City councilors Bur.	Bur.	Secretaries Leg.	Mayor Leg.	Exec.	Bur.	Exec.	Bur.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\hat{\beta}$	−1.217**	0.003	−0.268	−0.373	3.686	−0.134*	−0.165**	−0.055	−0.113*
	(0.510)	(0.074)	(0.372)	(3.488)	(3.204)	(0.073)	(0.066)	(0.082)	(0.067)
N	539	858	858	232	234	538	853	538	854
R^2	0.147	0.081	0.090	0.281	0.242	0.245	0.138	0.191	0.125

Models 1-9 control for randomization pair fixed effects
Standard errors clustered at the municipality level for models 1-9. *p<0.1; **p<0.05; ***p<0.01

under which providing information to citizens leads to improved governance outcomes, we know little about whether and how providing information to politicians improves governance.

To address this question, I partnered with the state audit court of Rio Grande do Norte (TCE-RN), in the Brazilian northeast, to implement an experimental study of information and local government accountability. In a random half of all but 5 of the state's municipalities, we treated local politicians (mayor, city councilors, and secretaries of key areas) with a scorecard containing information about the composition and performance of the local bureaucracy, as well as reminders about the main rules for hiring municipal employees and their responsibility in ensuring compliance with rules and alignment between spending and improvements in the quality of public services. The main outcome of the experiment is hiring irregularities, as measured through monthly payroll data obtained by TCE-RN. Intermediate outcomes are knowledge, self-efficacy, and perceptions and behaviors of accountability pressure among local politicians and bureaucrats, as measured through an original face-to-face survey of bureaucrats and an online survey of politicians.

Data collection is still ongoing. Complete payroll data for March, which is when I pre-specified hiring irregularities would be measured, will only be available in July. The survey of politicians was launched later than expected and so far only about 32% of the target population have responded. Results presented here leverage payroll data for February (when some of the temporary contracts for the year have not been created yet), politician survey data to date, and the complete survey of

bureaucrats. These preliminary results suggest that the treatment reduced the share of temporary contracts by about 0.17 standard deviations in average ($p < 0.05$)). This result however is still to be confirmed with final payroll data for the month of March, and to be adjusted for multiple testing.

I leverage original surveys of bureaucrats and politicians to explore some of the treatment's mechanisms, in line with the pre-analysis plan. Using preliminary survey data, the treatment shows an average increase of 0.14 standard deviations in politicians' knowledge about hiring rules ($p < 0.05$). The treatment *reduced* perceptions of accountability pressure from the state audit court, which sent the scorecards, by about 0.18 standard deviations ($p < 0.05$). This could be due to politicians reacting to an increased sense of oversight and surveillance by questioning the legitimacy of the court and/or blaming them for the state of affairs. Finally, some of the results suggest the treatment may have changed relationships between executive actors (mayors, secretaries, and bureaucrats) and city councilors, in a way that is consistent with executive actors engaging in blame avoidance.

Final results are expected by August, when I will have access to complete payroll and survey data, and thus be able to do more concluding analyses about the effects of the intervention. By then I will also include results of the conjoint experiment, randomization inference p-values and the multiple testing corrections for the tests of hiring irregularities presented in Table 2. Final results are expected to contribute to the emerging experimental literature on the dynamics of within-government accountability, as well as to policy debates about the effectiveness of information interventions.

References

Abranches, S. (1988). Presidencialismo de coalizão: o dilema institucional brasileiro. *Dados 31*(1), 5–38.

Adida, C., J. Gottlieb, E. Kramon, and G. McClendon (2016). Salience and coordination jointly moderate the impact of information on vote choice: Experimental evidence from benin. *Working Paper*.

Adida, C., J. Gottlieb, E. Kramon, and G. McClendon (2017). Overcoming or reinforcing coethnic preferences? an experiment on information and ethnic voting. *Working Paper*.

Akhtari, M., D. Moreira, and L. Trucco (2018). Political Turnover, Bureaucratic Turnover, and the Quality of Public Services. *Unpublished manuscript*.

Arias, E., H. Larreguy, J. Marshall, and P. Querubin (2017). Priors rule: When do malfeasance revelations help or hurt incumbent parties? *Working paper*.

Arretche, M. (2004). Federalismo e políticas sociais no brasil: problemas de coordenação e autonomia. *São Paulo em perspectiva 18*(2), 17–26.

Avis, E., C. Ferraz, and F. Finan (2017). Do government audits reduce corruption? estimating the impacts of exposing corrupt politicians. *National Bureau of Economic Research*.

Barrios, T. (2014). Optimal stratification in randomized experiments. *Mimeo*.

Boas, T. C., F. D. Hidalgo, and M. A. Melo (2018). Norms versus action: Voting against malfeasance in brazil.

Bobonis, G. J., L. R. Cámara Fuertes, and R. Schwabe (2016). Monitoring corruptible politicians. *American Economic Review 106*(8), 2371–2405.

Buntaine, M. T., R. Jablonski, D. L. Nielson, and P. M. Pickering (2018). Sms texts on corruption help ugandan voters hold elected councillors accountable at the polls. *Proceedings of the National Academy of Sciences*, 201722306.

Colonnelli, E., M. Prem, and E. Teso (2018). Patronage in the allocation of public sector jobs. *Working paper*.

De la O, A. and F. M. García (2015). Can intrastate accountability reduce local capture? results from a field experiment in mexico. *Working paper*.

Gottlieb, J. (2016). Greater expectations: A field experiment to improve accountability in mali. *American Journal of Political Science 60*(1), 143–157.

Hidalgo, F. D., J. Canello, and R. Lima-de Oliveira (2016). Can politicians police themselves? natural experimental evidence from brazil's audit courts. *Comparative Political Studies 49*(13), 1739–1773.

Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Lagunes, P. (2017). Guardians of accountability: A eld experiment on corruption and ine ciency in local public work. *IGC working paper*.

Leal, V. N. (1948). *Coronelismo, enxada e voto: o município e o regime representativo no Brasil*. Editora Companhia das Letras.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more t in experiments. *Journal of development Economics 99*(2), 210–221.

Melo, M. A., C. Pereira, and C. M. Figueiredo (2009). Political and institutional checks on corruption: Explaining the performance of brazilian audit institutions. *Comparative Political Studies 42*(9), 1217–1244.

O'Donnell, G. A. (1998). Horizontal accountability in new democracies. *Journal of democracy 9*(3), 112–126.

Power, T. J. (2010). Optimism, pessimism, and coalitional presidentialism: Debating the institutional design of brazilian democracy. *Bulletin of Latin American Research 29*(1), 18–33.

Praça, S. and M. M. Taylor (2014). Inching toward accountability: The evolution of brazil's anticorruption institutions, 1985–2010. *Latin American Politics and Society 56*(2), 27–48.

Raffler, P. (2018). Does political oversight of the bureaucracy increase accountability? field experimental evidence from an electoral autocracy.

Sakai, J. and N. Paiva (2016). Quem são os conselheiros dos tribunais de contas. *Transparência Brasil, São Paulo*.

TCE-RN (2018). Conhecendo o tce-rn.

Toral, G. (2019a). The benefits of patronage: How the political appointment of bureaucrats can enhance effectiveness and accountability. *Working paper*.

Toral, G. (2019b). Political bureaucratic cycles: How politicians manipulate public employment ahead of elections, and why that matters for human development. *Working paper*.

Westfall, P. H. and S. S. Young (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. Wiley-Interscience.

Zamboni, Y. and S. Litschig (2018). Audit risk and rent extraction: Evidence from a randomized evaluation in brazil. *Journal of Development Economics 134*, 133–149.

Appendices

A Location of the experiment 26

B Sample scorecard 28

C Additional details on hypotheses and variables used to test them 29

C.1 Increases in accountability inputs 29

C.2 Increases in accountability outputs 31

C.3 Increases in accountability outcomes 33

C.4 Heterogeneous treatment effects 35

D Survey instruments 37

E Research assistants for field survey of bureaucrats 37

F Ethical considerations 38

G Balance in pre-treatment covariates 41

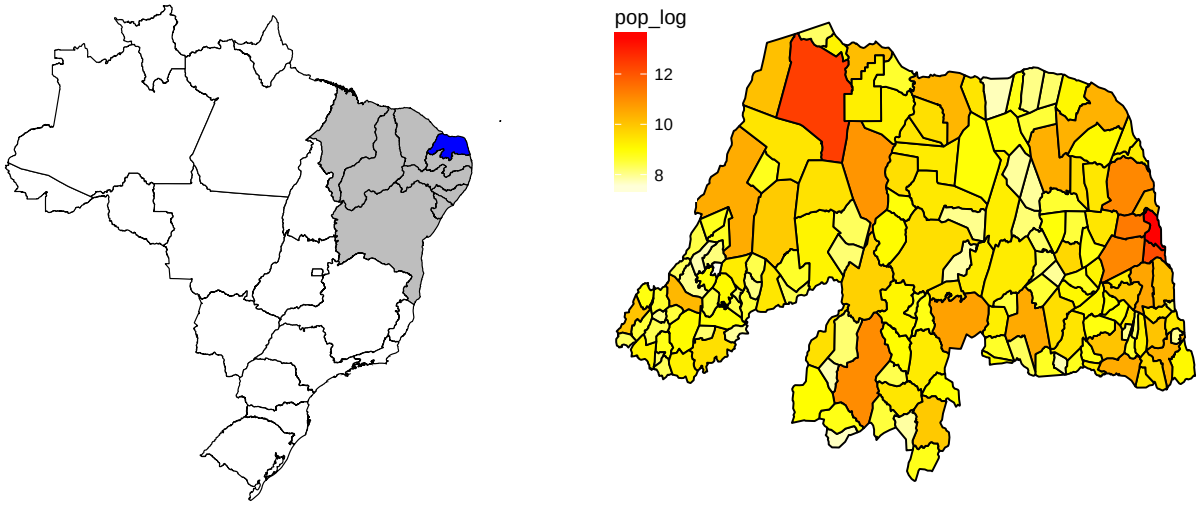
H Additional figures 42

I Additional tables 44

J Deviations from the pre-analysis plan 47

A Location of the experiment

Figure A.1: Location of the field experiment



(a) Brazil
(in grey, the Northeast region; in blue, the state of Rio Grande do Norte)

(b) Rio Grande do Norte
(colors correspond to the logged population of each municipality)

Figure A.2: Municipalities in the treatment group (in blue), in the control group (in white), and excluded from the experimental sample (in grey)

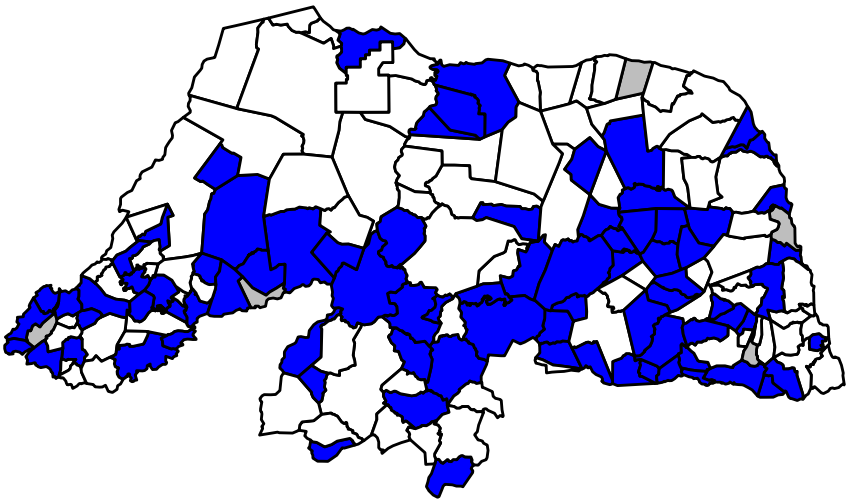
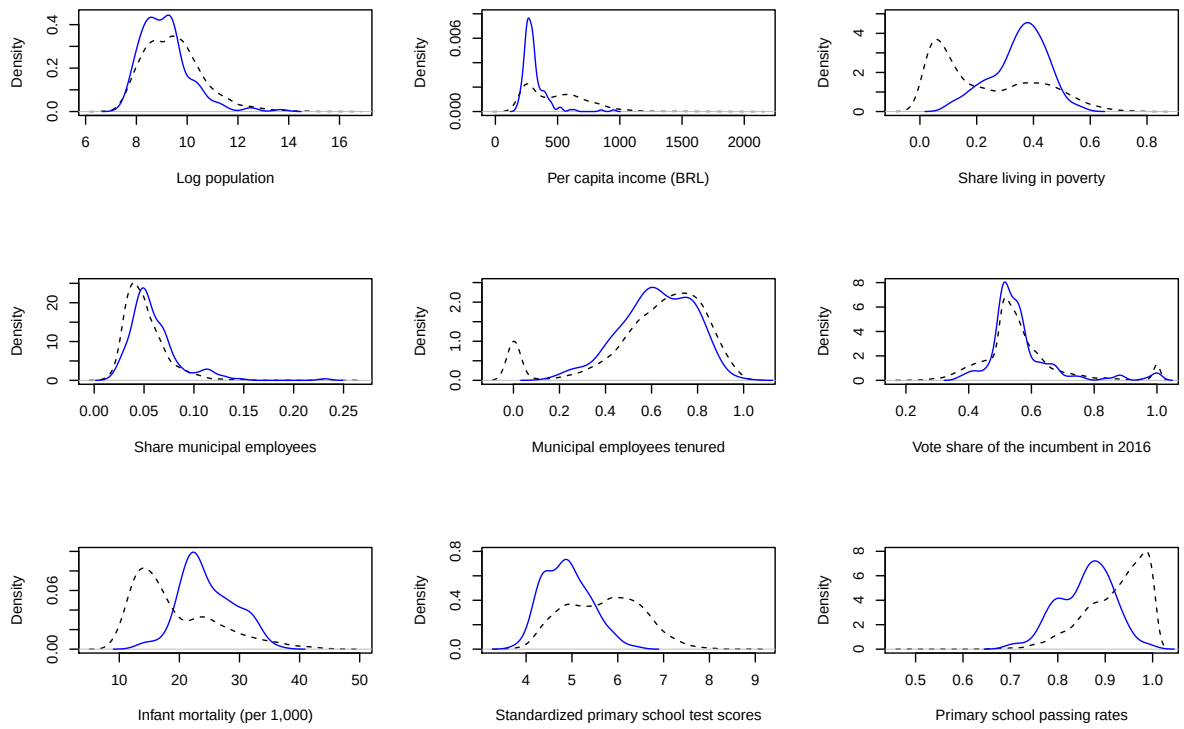


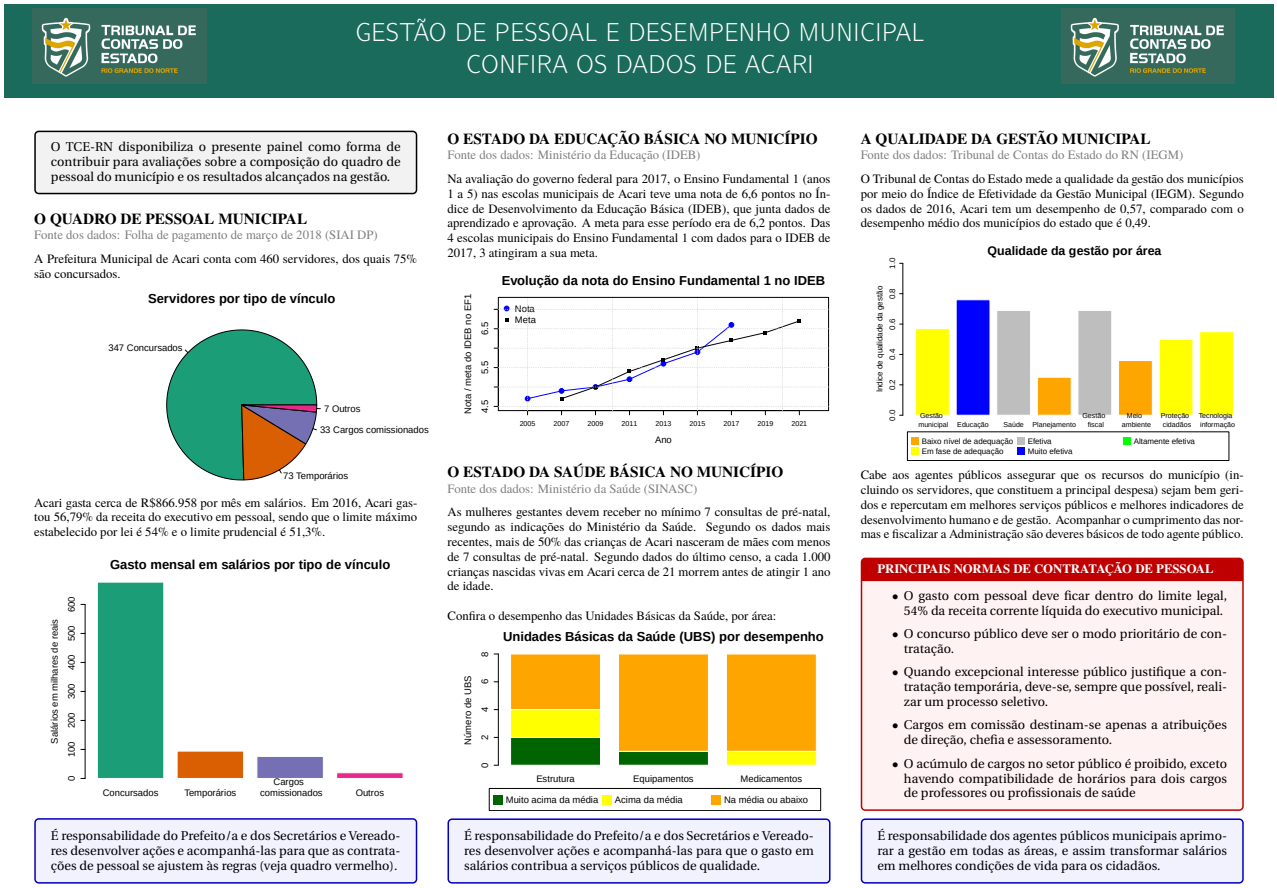
Figure A.3: Statistics for municipalities in Rio Grande do Norte (continues, blue line), compared to all municipalities in Brazil (dashed, black line)



Data are from IBGE (Brazil's official statistics institute) and the Ministry of Education

B Sample scorecard

The scorecard was designed in consultation with the implementing partner, TCE-RN. Scorecards were printed in A3-sized paper and delivered by registered mail in November, and through electronic means again in January 2019. The 2019 scorecards were the same as those sent in November, except for the data on IEGM which was updated.



C Additional details on hypotheses and variables used to test them

Below each hypothesis, this section details what survey items are used to test it, as per the pre-analysis plan.

C.1 Increases in accountability inputs

H1: Sending the scorecard improves politicians' knowledge about the formal rules that hiring decisions need to follow. The scorecard will highlight some of the most important formal rules in hiring and increase their salience, so I expect it will increase knowledge among officials. The key assumption here is that politicians do not know these rules already. In my interviews I have found knowledge about rules is mixed.

- *In what cases is it legal to hire workers under a temporary contract? (Mark all that apply)*
- *In what cases is it legal to hire workers under a position of trust [cargo comissionado]? (Mark all that apply)*
- *What is the maximum share of its income that the executive municipal government can spend on personnel?*

H2: Sending the scorecard improves politicians' knowledge about the municipality's compliance with formal rules on hiring. The scorecard will also contain specific information on how the municipality performs with respect to those rules, so knowledge about compliance should also increase. Here the assumption is that politicians lack this knowledge, which is probably more true for legislators than for secretaries.

- *What percent of its revenue does the municipal government spend on personnel?*
- *What is the proportion of the municipality's workers who are hired under a temporary contract?*
- *What is the proportion of the municipality's workers who are hired under a position of trust?*

H3: Sending the scorecard improves politicians' knowledge about the performance of the municipality's bureaucracy. In a similar vein, I expect the scorecard to increase knowledge about the performance of the local bureaucracy. Again I expect the effect to be stronger among city councilors (who are less exposed to bureaucratic management and the reporting mechanisms imposed by the federal government).

- *What score did the municipality receive in the primary school quality index IDEB, as per the results published in September?*
- *Did the municipality reach its target for the primary school quality index IDEB, as per the results published in September?*
- *What score did the municipality get in TCE-RN's municipal management quality index (IEGM)?*

For all these knowledge questions it is likely that some politicians will not respond. The survey instrument encourages them to give an approximate answer, even if they do not know the exact figure. To deal with potential missing data, I will use as outcomes both politician answers and whether they decide to respond even if it is with an approximate answer.

H4: Sending the scorecard increases politicians' perceptions of accountability pressure from the state audit court, TCE-RN. The scorecard is expected not only to provide information but also to increase subjective feelings of accountability pressures from the state audit court, since the state audit court can impose penalties for irregularities detected.

- *"I trust TCE-RN" (agreement on a 1-4 scale)*
- *"TCE-RN is concerned with improving municipal management" (agreement on a 1-4 scale)*
- *"TCE-RN knows the reality of this municipality" (agreement on a 1-4 scale)*
- *"TCE-RN detects irregularities that take place in the management of this municipality" (agreement on a 1-4 scale)*

H5: Sending the scorecard increases politicians' sense of self-efficacy in the control of irregularities in hiring. The scorecard will highlight politicians' responsibilities in shaping the local bureaucracy's

compliance, and hence I expect it to increase their sense of self-efficacy, especially among city councilors. By self-efficacy I refer to politicians' beliefs that they are responsible for and capable of controlling irregularities in hiring.

- *Rank the following actors according to who is most responsible in controlling irregularities in hiring: mayor, secretaries, city councilors, bureaucrats, citizens, the Public Prosecutor's Office, the state audit court.*

H6: Sending the scorecard increases politicians' sense of self-efficacy in overseeing the bureaucracy and fostering its performance. The scorecard will also highlight politicians' responsibilities in fostering the performance of the municipal bureaucracy, and hence I expect it to increase their sense of self-efficacy in this respect.

- *Rank the following actors according to who is most responsible for improving the quality of public services: mayor, secretaries, city councilors, bureaucrats, citizens, the Public Prosecutor's Office, the state audit court.*

C.2 Increases in accountability outputs

If the treatment increases accountability inputs (knowledge, self-efficacy, and perceived accountability pressure) I hypothesize it will increase – to a lesser extent – accountability outputs or behaviors among politicians.

H7: Sending the scorecard improves politicians' understanding that the use of bureaucratic appointments as patronage should be limited. I expect the treatment to improve attitudes about the use of patronage in municipal bureaucracies.

- Conjoint experiment where respondents see pairs of hypothetical municipal employees, with attributes randomly varied. To test this hypothesis, I will look at the average marginal component effect of an employee having political connections on it being chosen as more likely to be appointed for a position of street level management (such as school director or clinic manager).

H8: Sending the scorecard increases legislators' accountability pressure on the executive. By informing and empowering legislators, who usually operate in environments with less information, I expect the treatment will foster the accountability pressure of city councilors on the executive.

- *How many times in the past 3 months have you met with a city councilor?* (for politicians in the executive)
- *How many times in the past 3 months have you met with a city councilor?* (for bureaucrats)
- *How many times in the past 3 months have you been contacted by a city councilor?* (for bureaucrats)
- *How many times in the past 3 months have you met with a secretary?* (for politicians in the legislative)
- *How many times in the past 3 months have you met with the mayor?* (for politicians in the legislative)
- *"City councilors are well informed about the reality of public services in your municipality" (agreement on a 1-4 scale)* (for politicians in the executive and bureaucrats)
- *"City councilors work to improve public services for the all" (agreement on a 1-4 scale)* (for politicians in the executive and bureaucrats)

H9: Sending the scorecard increases the accountability pressure of politicians in the executive (the mayor and secretaries) on the bureaucracy. I hypothesize the treatment will increase monitoring and oversight of the bureaucracy by the political leadership.

- *How many times in the past 3 months have you visited a municipal school?* (for politicians in the executive)
- *How many times in the past 3 months have you visited a municipal clinic?* (for politicians in the executive)
- *How many times in the past 3 months have you visited a municipal social assistance center?* (for politicians in the executive)

- *How many times in the past 3 months have you met with secretaries?* (for the mayor only). The idea here is that the mayor may choose to put pressure on the bureaucracy via their secretaries.
- *How many times in the past 3 months have you met with the mayor?* (for secretaries only)
- *How many times in the past 3 months have you met the secretary in your area?* (for bureaucrats only)
- *How many times in the past 3 months have you met with technicians from the secretariat of your area?* (for bureaucrats only)
- *How many times in the past 3 months have you met the mayor?* (for bureaucrats only)
- *How many times in the past 3 months have you been contacted by the mayor?* (for bureaucrats only)
- *How many times in the past 3 months have you been contacted by the secretary in your area?* (for bureaucrats only)
- *"The mayor and [education / healthcare / social assistance] professionals have the same priorities in relation to [the schools / the clinics / the social assistance centers]" (agreement on a 1-4 scale)* (for bureaucrats only)
- *"The secretary of [education / healthcare / social assistance] knows the reality of this [school / clinic / social assistance center]" (agreement on a 1-4 scale)* (for bureaucrats only)
- *"The secretariat of [education / healthcare / social assistance] holds this [school / clinic / social assistance center] accountable for its results" (agreement on a 1-4 scale)* (for bureaucrats only)

C.3 Increases in accountability outcomes

Last, I hypothesize the treatment will increase development outcomes, in two areas – decreases in hiring irregularities (which is my primary outcome) and increases in bureaucratic effectiveness (effects which I expect to be weaker and to take longer to show in the data – I think of these as

downstream outcomes). My prior is that the experiment will have weak or no effects in accountability outcomes, since the treatment is relatively weak, the experiment has a small sample size, and these outcomes are harder to influence.

H10: Sending the scorecard decreases hiring irregularities. Through some or all of the causal chains hypothesized above, I hypothesize the scorecard decreases hiring irregularities, as measured through administrative payroll data:

- Percent of municipal employees who are not tenured. This is the experiment's main outcome, which was used in the pairing algorithm. To examine the drivers of any potential treatment effects, I will also look at the following variables that can be seen as components of the share of non-tenured employees:
  - Percent of municipal employees under temporary contracts.
  - Percent of municipal employees in positions of trust.
  - Logged total number of municipal employees.
- Logged total amount spent in salaries.
- Share of municipal income spent in personnel.
- Number of municipal employees who have more than one or two positions in the public sector.

H11: Sending the scorecard improves bureaucratic performance. These are some measurable downstream outcomes that the experiment might impact.

- Municipal primary school average grade (Portuguese and Mathematics), as measured by the Ministry of Education's IDEB in late 2019 (data to be released in 2020).
- Municipal primary school average passing rate, as measured by the Ministry of Education's IDEB in late 2019 (data to be released in 2020).
- Number of pre-natal check-ups, as measured by the Ministry of Healthcare's SIAB during the first semester of 2019 (possibly not released til 2020).
- Number of healthcare home visits, as measured by the Ministry of Healthcare's SIAB during the first semester of 2019 (possibly not released til 2020).

- Number of families who benefit from the social assistance program PAIF during August 2019 (data released in 2020), as measured by the Ministry of Social Development's Social Assistance Census.

C.4 Heterogeneous treatment effects

With only 162 municipalities in the experimental sample, the experiment is severely under-powered to examine heterogeneity in treatment effects at the municipality level – interactions would have to be very significant for the experiment to be able to pick them up. I hypothesize nonetheless some potential sources of heterogeneity which may have a large effect.

H12: Increases in accountability outputs and outcomes are particularly likely where the mayor has a low level of support in the legislature. Mayors who lack a strong support basis in the chamber are arguably more vulnerable and hence more sensitive to manipulations of accountability inputs. The level of support in the legislature will be measured with the average perception about politicians of the share of legislators who support the mayor. The hypothesis will be tested by interacting the treatment indicator with an indicator for whether that level of support is below the state-wide median.

H13: Increases in accountability outputs and outcomes are particularly likely where the mayor is in their first term and is thus eligible for re-election. Mayors who are eligible for re-election are arguably more responsive to increases in accountability inputs. The hypothesis will be tested by interacting the treatment indicator with an indicator for whether the mayor is in their first term.

H14: Increases in accountability outputs and outcomes are particularly likely where the share of municipal employees is lower (below the median). Municipal governments that have less employees under tenured contracts have more room for maneuver to decrease patronage and cut personnel expenses. They may also be more able to put pressure onto the bureaucracy to increase performance. This hypothesis will be tested by interacting the treatment indicator with an indicator for whether the share of tenured employees (at baseline) is below the state-wide median.

On the other hand, some individual-level heterogeneity might be detected at the individual level.

H15: Increases in accountability inputs are larger among legislators. I hypothesize city councilors

to be more sensitive to the information treatment given their lower baseline levels of information, and self-efficacy.

H16: Increases in accountability inputs are larger among opposition legislators. I hypothesize opposition city councilors (those who do not declare supporting the mayor) to be more sensitive to the information treatment given their even lower baseline levels of information. I will test this hypothesis by interacting

H17: Increases in accountability pressure on bureaucrats are larger among those who have closer ties to the executive. I hypothesize municipal governments will put more accountability pressure on bureaucrats that have closer ties to politicl leadership (mayor, secretaries).

D Survey instruments

The survey instruments were pre-registered with the rest of the design, and can be consulted online:

- Survey of politicians in Portuguese and in English
- Survey of school directors in Portuguese and in English
- Survey of clinic managers in Portuguese and in English
- Survey of social assistance center coordinators in Portuguese and in English

E Research assistants for field survey of bureaucrats

The following 23 people provided excellent research assistance for the implementation of the field survey in Rio Grande do Norte: Jenair Alves, Marcos Aurélio Freire da Silva Júnior, Francymonni Yasmim Marques de Melo, Karoline de Oliveira, Raiany Juliete da Sila, Aline Juliete de Abreu Feliciano, Pedro Henrique Correia do Nascimento Oliveira, Ana Vitória Araújo Fernandes, Jaedson Gomes dos Santos, Ana Beatriz Germano Barroca, Renata Lima de Morais, Myleyde Dayane Pereira da Silva, Marina Rotenberg, Filipe Ramos Pinheiro, Daniele Vitória Lima da Silva, Elvira Gomes Santos, Matheus Oliveira de Santana, Magda Emanuele Lima da Silva, Ayanne Marília Sousa da Silva, Júlio César Nascimento, Lidiane Freire de Jesús, André Silva, and Pâmela Kaissa Fernandes Lopes.

F Ethical considerations

The study was reviewed and approved by the Committee on the Use of Humans as Experimental Subjects at MIT. In the paragraphs below, I discuss potential ethical concerns related to the experiment, and argue why the intervention involves no more than minimal risk for subjects.

Informed consent. Human subjects in the experiment are the public officials who received the scorecards – namely, mayors and city councilors, who are elected, and secretaries, who are appointed by mayors. The experiment did not seek their informed consent for two main reasons. First, making them aware that their receiving the scorecard is part of an experimental study could have severely jeopardized the validity of the experiment, since they would most likely have disregarded the information. Second, politicians constantly receive information from multiple government agencies, including the state audit court. With regards to the surveys used to measure some of the experiment's outcomes, survey respondents (mayors, city councilors, secretaries, and bureaucratic managers) will go through standard informed consent procedures before deciding to participate in the survey.

Deception. Human subjects in the experiment – public officials of municipalities in the treatment group – were not be deceived. They received information that reflected real, up-to-date, high-quality administrative data. Data sources were included in the scorecard. Datasets used to build the scorecard are all publicly available government data, and the municipality's own payroll data which they submit monthly to TCE-RN.

Researcher involvement. The state audit court of Rio Grande do Norte regularly sends municipal politicians information and persuasion campaigns to try and constrain corruption, foster compliance, and increase efficiency in the use of public funds. Nonetheless, the court lacks a rigorous empirical base for understanding whether and how these campaigns influence municipal finances and governance. The information campaign in the experiment differs from other campaigns in three main aspects, as a result of the my involvement. First, this campaign introduces an experimental design, where a random half instead of all the municipalities in the state will be exposed to this campaign. Second, the content and format of the campaign was designed joining data from multiple sources (including publicly available administrative data, and payroll data submitted by the municipalities to the court) to create scorecards that were printed in high-quality, color posters instead of usual black and white letters. Third, the campaign was accompanied by surveys of bureaucrats and politicians to better understand the effects of the information.

Risks. The experimental intervention consisted merely of an information campaign, leveraging information that is already public or known to politicians. This makes the intervention weak by design. The intervention was expected to have a positive effect (if any), by decreasing irregularities in hiring, freeing up municipal government resources, improving local governance, and fostering the performance of the local bureaucracy. Municipal elections are scheduled for almost two years after the intervention (in October of 2020), and state and federal elections will happen two years later (in October of 2022). Therefore, the intervention is very unlikely to affect elections. I foresee three potential adverse effects for some individuals. In all three cases, net benefits for citizens in the municipality are likely to outweigh the potential adverse effects for some politicians or bureaucrats.

- Some municipal employees could perhaps see their jobs (contract, conditions, etc.) impacted by the intervention, if it makes treated politicians in the executive change their hiring decisions. Because of the design of the scorecard, the jobs that are most likely to be affected are those that were created as patronage or clientelistic positions -– in that sense, while that particular individual may be negatively affected as a result of the intervention, the municipality as a whole would benefit through a more efficient use of resources.

- By providing information to city councilors (that is, politicians outside the executive government), and potentially increasing their knowledge, self-efficacy and capacity to exercise their accountability role vis-à-vis the municipal government, the intervention could intensify accountability pressures among the legislative and the executive, and/or between politicians in power and in the opposition. To the extent the intervention does intensify accountability pressures from the legislative on the executive (which may bring costs to politicians in the executive), most theories in political science predict these pressures to have a positive effect on local governance, government performance, and citizen welfare.

- Finally, both the intervention itself and the surveys may take time away from bureaucrats and politicians that they would better invest in other tasks like serving citizens. Surveys are however designed to take little time (15 minutes for politicians and 30 for bureaucrats). More importantly, these time investments are expected to yield benefits in terms of knowledge (for academics and policymakers, including TCE-RN). Finally, survey participants often reported having enjoyed participating in the research and valuing the data collection effort.

Beneficience. Beyond the intervention's expected benefits and the strategies described above to minimize risks, the study includes a number of actions that are expected to benefit the state

where the experiment is being implemented. First, I will deliver at least two training sessions in Natal, which are expected to increase the capacity of local researchers and policymakers for evaluating public policies. These consist of a two-hour training introducing state auditors to the logic of randomized control trials (delivered in December of 2018), and a four-day training on experiments for public policy (to be delivered at the federal university in July of 2019 with graduate students and municipal secretaries of education). If treatment effects and the learning experience are beneficial, this is expected to be the first of many experiments implemented by the state audit court to widen the evidence base for their accountability work. Second, I will present the results to TCE-RN through an in-person seminar and a policy-oriented report in Portuguese, so they can use the results internally to orient their work as well as to launch discussions on their influence over local governance. Third and last, the project trained 23 enumerators who are students or recent graduates of the local federal university, and provided them with skills and experience in surveys that can later be leveraged for other local research projects.

Anonymity of the survey data. All survey data is being kept confidential and stored in secured hard drives. Only aggregate results will be published, and no piece of information that can identify respondents will be made publicly available.

G Balance in pre-treatment covariates

Table A1: Balance in pre-treatment covariates between treated and untreated municipalities

	mean treated	mean control	p value t test	p value KS test
total_servidores_log	6.10	6.18	0.47	0.72
populacao_2016_log	9.09	9.23	0.34	0.52
populacao_rural_pc	0.35	0.35	0.92	0.59
pc_pobres	35.51	33.98	0.33	0.20
pc_extremamente_pobres	18.41	16.18	0.07	0.10
renda_percapita	300.65	306.38	0.67	0.88
mayor_1620_voteshare_16	0.56	0.55	0.68	0.61
mayor_1216_voteshare_12	0.57	0.57	0.94	0.89
mayor_reelected_2016	0.25	0.21	0.52	
mayor_reelected_2012	0.27	0.29	0.82	
runnerup_1620_wasincumbent	0.20	0.17	0.74	
margin_of_victory_2016	0.14	0.14	0.90	0.93
margin_of_victory_2012	0.16	0.17	0.79	0.83
herfindahl_08	0.53	0.52	0.57	0.72
herfindahl_12	0.52	0.51	0.60	0.76
herfindahl_16	0.50	0.49	0.61	0.74
share_enrolment_mun_2016	0.71	0.70	0.80	0.66
household_monthly_pc_income_2010	276.46	279.59	0.76	0.98
internet_2014	0.59	0.52	0.43	
radio_am_2012	0.14	0.13	0.82	
radio_fm_2012	0.21	0.21	0.94	
gini	0.50	0.49	0.16	0.15
desocupacao_18a24	18.42	19.02	0.67	0.85
bf_families_2016_log	7.15	7.22	0.65	0.38
criancas_0a5_foradaescola	49.34	50.65	0.32	0.69
esperancadevida_aonascer	70.68	71.01	0.21	0.21
mortalidade_infantil	24.93	24.01	0.22	0.20
mortalidade_ate5anos	26.79	25.81	0.22	0.22
idhm	0.61	0.61	0.88	0.98
idhm_renda	0.58	0.58	0.67	0.85
idhm_longevidade	0.76	0.77	0.21	0.19
idhm_educacao	0.52	0.52	0.68	0.85
receita	44054532.55	40242106.67	0.86	0.61
despesa	42986322.58	38770666.05	0.84	0.49
fiscal_balance	1068209.97	1471440.62	0.59	0.56
iegm_nota	0.49	0.49	0.68	0.83
ieduc_nota	0.52	0.54	0.34	0.89
isaude_nota	0.65	0.62	0.20	0.09
iplanej_nota	0.26	0.25	0.54	0.28
ifiscal_nota	0.65	0.68	0.13	0.17
iamb_nota	0.30	0.34	0.23	0.60
icidade_nota	0.34	0.34	0.93	0.09
igovti_nota	0.45	0.44	0.69	0.53
total_servidores_sq_log	12.20	12.36	0.47	0.72
populacao_2016_sq_log	18.17	18.46	0.34	0.52

H Additional figures

Figure H.4: Largest 17 municipalities, excluded from the sampling frame for the survey of street level bureaucrats

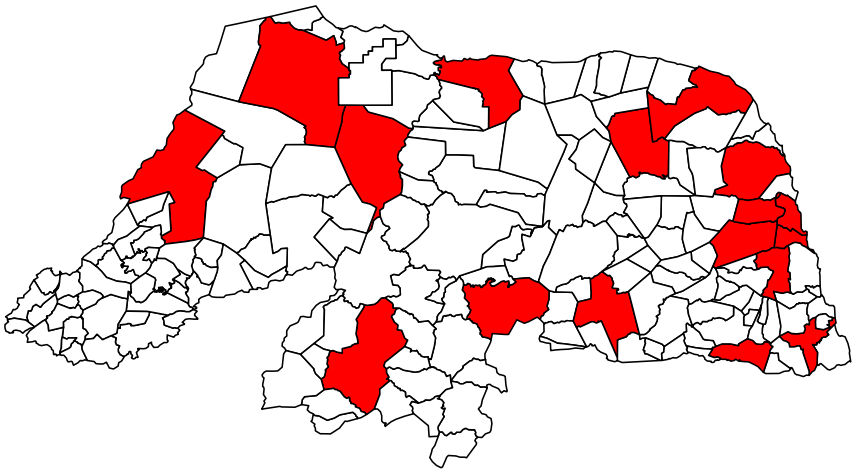


Figure H.5: Number of face-to-face surveys of bureaucrats done by municipality (municipalities in white are excluded from the survey)

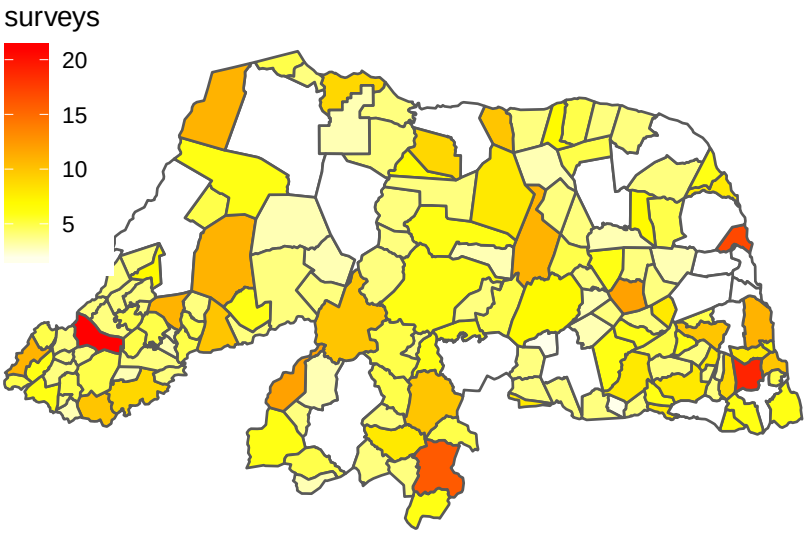
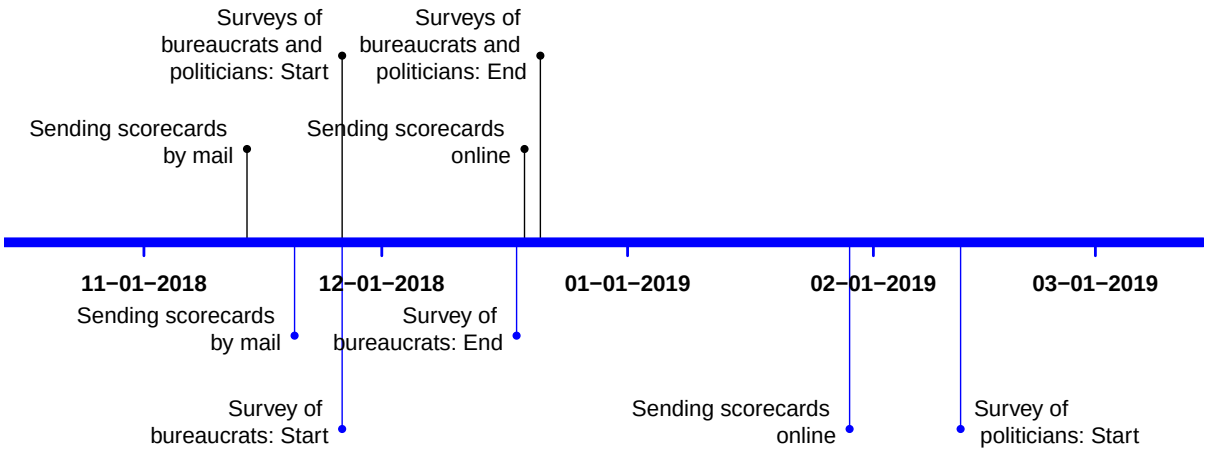


Figure H.6: Planned (above) versus actual (below) milestones for implementation



I Additional tables

Table A2: Treatment effect on politicians' knowledge about the municipality's compliance with rules (hypothesis 2), measured through an ongoing online survey

	Dependent variable: Knowledge about the municipality's compliance with rules...				
	Personnel spending	% workers temporary	% workers positions of trust	PCA index	Average score
	(1)	(2)	(3)	(4)	(5)
$\hat{\beta}$:	0.062	−0.013	−0.032*	−0.016	0.001
	(0.042)	(0.018)	(0.017)	(0.197)	(0.016)
N	364	378	391	285	463
R^2	0.509	0.309	0.355	0.340	0.314

Knowledge measured as distance between respondent's guess and municipality's performance
All regressions control for randomization pair fixed effects
Standard errors clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01

Table A3: Treatment effect on politicians' knowledge about the municipality's performance (hypothesis 3), measured through an ongoing online survey

	Dependent variable: Knowledge about municipal performance			
	IDEB score	IDEB target met	PCA index	Average score
	(1)	(2)	(3)	(4)
$\hat{\beta}$:	0.011	0.008	0.029	0.009
	(0.031)	(0.037)	(0.092)	(0.030)
N	854	854	854	854
R^2	0.164	0.169	0.170	0.170

Knowledge measured as whether respondent's guess was correct.
All regressions control for randomization pair fixed effects
Standard errors clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01

Table A4: Treatment effect on politicians' perceptions of accountability pressure from the state prosecutors office (not pre-registered)

	Dependent variable: Agreement with statements about MP-RN				
	I trust MP-RN	MP-RN knows this municipality	MP-RN detects irregularities in this municipality	PCA index	Average score
	(1)	(2)	(3)	(4)	(5)
$\hat{\beta}$: treatment	-0.048	-0.142^{**}	-0.156^{***}	-0.248^{***}	-0.115^{***}
	(0.048)	(0.060)	(0.060)	(0.094)	(0.043)
N	771	771	771	771	771
R^2	0.122	0.114	0.147	0.132	0.133

All regressions control for randomization pair fixed effects
Standard errors clustered at the municipality level. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A5: Treatment effect on politicians' self-efficacy in the control of hiring irregularities (hypothesis 5) and in overseeing bureaucrats and improving their performance (hypothesis 6)

	Dependent variable: Self-efficacy in	
	controlling hiring irregularities	improving public services
	(1)	(2)
$\hat{\beta}$: treatment	-0.090	-0.002
	(0.118)	(0.099)
N	795	786
R^2	0.125	0.095

Self-efficacy measured as how respondents rank their position among a list of 7 actors
All regressions control for randomization pair fixed effects
Standard errors clustered at the municipality level. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A6: Treatment effect on executive politicians' accountability pressure on bureaucrats (hypothesis 9), as measured through the survey of politicians

	Dependent variable:				
	Visits by executive politicians to			Meetings of	
	schools	clinics	social assistance	mayor	secretaries
	schools	clinics	center	with secretaries	with mayor
	(1)	(2)	(3)	(4)	(5)
$\hat{\beta}$:	−0.736	−0.078	−0.735	13.304	0.603
	(0.940)	(1.180)	(0.954)	(54.028)	(1.204)
N	540	540	537	57	480
R^2	0.125	0.128	0.121	0.716	0.197

All regressions control for randomization pair fixed effects
Standard errors clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01

Table A7: Treatment effect on executive politicians' accountability pressure on bureaucrats (hypothesis 9), as measured through the survey of bureaucrats

	Dependent variable:							
	Meetings with			Contacts from		Agreement with		
	secretary	technicians	mayor	mayor	secretary	mayor & professionals are aligned	secretary knows this unit	secretariat holds us accountable
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\hat{\beta}$: treatment	−0.961	0.222	0.100	0.775	2.377	−0.033	0.020	0.035
	(0.895)	(0.602)	(0.375)	(0.995)	(2.314)	(0.060)	(0.039)	(0.028)
N	847	858	858	858	848	856	845	854
R^2	0.153	0.097	0.127	0.147	0.164	0.158	0.149	0.120

All regressions control for randomization pair fixed effects
Standard errors clustered at the municipality level. *p<0.1; **p<0.05; ***p<0.01

J Deviations from the pre-analysis plan

After review by TCE-RN, two survey items were dropped from the final version of the survey that was fielded with politicians:

- *"TCE-RN has a positive influence the management of this municipality" (agreement on a 1-4 scale)* – pre-specified as linked to hypothesis 4.
- *"City councilors make inquiries about the state of public services in the municipality" (agreement on a 1-4 scale)* – pre-specified as linked to hypothesis 8.

For testing hypothesis 3 (namely that the scorecard would improve politicians' knowledge of performance of the bureaucracy) I pre-registered a survey item asking them about the municipality's performance in TCE-RN's governance index IEGM. While this item was included in the survey, measuring the effect of treatment on these responses faces two main obstacles. First, the updated scorecards sent in February included data for the 2017 IEGM, while the original scorecards sent in November included data for the 2016 IEGM. The existence of updated IEGM data was the "excuse" used to re-treat politicians in treated municipalities, but as a result there is not one single IEGM score that I can compare politicians' responses to. Second, some politicians responded to the question about IEGM using the numeric scale (which goes from 0 to 1), while others used the letter scale (A, B+, B, C+, C). While none of these two obstacles is insurmountable, including this item in the experimental analyses would require taking a number of arbitrary coding decisions. I therefore choose to exclude it instead.

The pre-analysis plan specified a sampling of municipalities among the 150 largest in the state for the field survey of bureaucrats. Instead, I included all 150 municipalities in the field survey.

A number of pre-specified adjustments to data analysis are still to be implemented: incorporating Westfall-Young multiple testing corrections for hypotheses 10 and 11; applying weights to municipality-level regressions by the number of responses in each municipality; and the calculation of randomization inference p-values. These adjustments will be included in the paper once final outcome data is available.

Treatment effects on politicians beliefs about patronage (hypothesis 7) will be tested with the analysis of the conjoint experiment once data collection is finalized.

Results to the analyses of heterogeneous treatment effects are yet to be included in the Appendix.