Government Oversight and Inter-Institutional Legibility: Evidence from Colombia

Supporting Information

October 3, 2025

Contents

A1 Contextual Information

A1.1 Situating the Colombian case

We use cross-national indices to situate Colombia relative to other cases, focusing on two sets of indicators:

1. **Institutional features:**
   - Degree of decentralization:
     - **Regional Authority Index** (Hooghe et al., 2021): Measures the extent of authority exercised by regional governments across dimensions such as policy scope, fiscal autonomy, and representation. The scale ranges from 0 to 30. The values for Colombia are based on 2018 data.
     - **Division of Power Index** (Varieties of Democracy project): Captures the vertical dispersion of power in a country. The composite index ranges from 0 to 1 and is based on expert-coded assessments of institutional features. Subcomponents included:
       - **Local Offices Relative Power**: Assesses the relative power of municipal-level government institutions.
       - **Regional Offices Relative Power**: Assesses the relative power of regional (e.g., departmental) government institutions.
   - Strength of horizontal accountability institutions:
     - **Horizontal Accountability Index** (Varieties of Democracy project, available at `www.v-dem.net`): Measures the degree to which state institutions (e.g., courts, audit agencies, ombudsman) are capable of and willing to oversee and sanction government misconduct. It ranges from 0 (no accountability) to 1 (strong accountability).

2. **Outcome-level characteristics:** (i.e., outcome being monitored by the PGN in our intervention)
   - Transparency practices and level of perceived corruption:
     - **Rule of Law Index** (World Justice Project, available at `https://worldjusticeproject.org/rule-of-law-index/`: Evaluates the extent to which laws are clear, publicized, stable, and applied evenly. This index incorporates factors such as constraints on government power, absence of corruption, and openness of government. It ranges from 0 to 1. Factors included:
       - **Absence of Corruption**: Assesses the extent to which public officials in the executive, judicial, legislative, and security branches use public office for private gain. This includes bribery, improper influence by public or private interests, and misappropriation of public funds or other resources.
       - **Open Government**: Measures the openness of government, defined by the extent to which it shares information, empowers people with tools to hold the government accountable, and fosters citizen participation in public policy deliberations. This includes publicized laws and government data, right to information, civic participation, and complaint mechanisms.
     - **Corruption Perceptions Index (CPI)** (Transparency International, available at `https://www.transparency.org/en/cpi/2024`): Captures expert perceptions of public sector corruption on a scale from 0 (highly corrupt) to 100 (very clean).

In Table A1, we compare Colombia across three reference groups: all countries in the world, Latin America and the Caribbean (LAC), and upper-middle-income (UMI) countries.

Index	Mean	p25	p75	Colombia's Score	Percentile (All)	Percentile (LAC)	Percentile (UMI)
1. Regional Authority Index	10.66	2	17.5	15.01	73%	85%	67%
2. Division of Power Index	0.49	.28	.88	0.99	98%	100%	100%
2.a. Regional offices relative power	0.24	-1.01	1.39	2.46	97%	100%	100%
2.b. Local offices relative power	0.86	.4	1.8	2.27	92%	91%	93%
3. Horizontal Accountability Index	0.62	.39	.88	0.83	67%	71%	84%
4. Rule of Law Index	0.56	.45	.63	0.50	40%	38%	34%
4.a. Absence of Corruption Index	0.52	.38	.64	0.39	26%	28%	16%
4.b. Open Government Index	0.52	.42	.61	0.64	80%	90%	97%
5. Corruption Perceptions Index	43.34	29	56	39.00	49%	52%	54%

Note: All indices correspond to 2020, except for the RAI, which is based on 2018 data. Percentile values indicate Colombia's relative position; higher percentiles reflect better performance compared to other countries in the respective group. LAC refers to Latin America and the Caribbean, and UMI to upper-middle-income countries. The number of countries used for comparison varies by index: *Regional Authority Index*: 95 countries (27 LAC, 28 UMI); *Division of Power Index* and *Horizontal Accountability Index*—179 countries (25 LAC, 46 UMI); *Rule of Law Index*: 128 countries (30 LAC, 39 UMI); *Corruption Perceptions Index*: 180 countries (30 LAC, 49 UMI); *Local Offices Relative Power*—169 countries (23 LAC, 45 UMI); *Regional Offices Relative Power*: 137 countries (20 LAC, 38 UMI); *Absence of Corruption Index* and *Open Government Index*: 128 countries (30 LAC, 39 UMI).

Table A1: Cross-national contextualization of the Colombian case. Indices 1-3 measure the institutional environment. Indices 4-5 measure relevant outcomes.

On the one hand, looking at its institutional features, the table shows that Colombia is a highly decentralized country, ranking between the 73rd and 98th percentiles globally across multiple decentralization indices, suggesting both that the central government delegates a large number of governance activities to decentralized governmernts. This increases the scope of possible agency problems between central and decentralized entities. It also shows that Colombia's oversight institutions are meaningful, as it ranks in the upper half of countries on horizontal accountability. This means that the implementing partner in our intervention carries a relatively important weight among public sector institutions, thereby rendering our treatment a credible cost of overisght.

On the other hand, looking at the outcome the central govenrment in our empirical settign is interested in monitoring, i.e., transparency practices and corruption, we note that Colombia is close to the median on the Corruption Perceptions Index—neither exceptionally clean nor severely corrupt—and even lower on the Rule of Law Index's *Absence of Corruption* sub-index (26th percentile globally, 28th percentile in LAC, and 16th percentile among UMI countries). These scores indicate persistent uncertainty about actual transparency and rule-following by local officials. In contrast, Colombia ranks very high on the Rule of Law Index's *Open Government* sub-index (80th globally, 90th in LAC, 97th in UMI), which measures public access to government information, citizen engagement, and legal frameworks for transparency. This is particularly relevant, as it directly reflects the dimension of governance that our intervention targets and that the PGN seeks to monitor.

Taken together, these features place Colombia in a middle range: it is neither an extreme nor an outlier case but rather a representative setting where oversight dynamics are both relevant and observable. This combination of decentralization and the potential for agency problems in the relevant dimension (corruption/transparency) suggests that agency problems may be severe, but the strength of horizontal accountability institutions are counterveiling and suggest that Colombia is neither a hard nor an easy case for the efficacy of direct communication from a watchdog entity at the central level.

A1.2 PGN oversight of public sector entities

We draw on two sources of data to understand the scope and targeting of PGN's oversight of public sector entities. These data sources are as follows:

- The *procesos disciplinarios elección popular* dataset from the PGN records the **elected officials investigated by the PGN** between 2016 and 2019 (prior to our intervention). We observe when a case was initiated and the progress of the case. Obviously, only a subset of public-sector entities including elected officials, namely the executive and legislative arms of the (1) national government; (2) departmental governments; and (3) municipal governments. In total, there are 2,259 entities in our experimental sample with elected principals, totaling 34.5% of the entities in the main sample. This helps us to understand the variation in the targeting of investigations ($\rho(r)$) in our framework.
- The *Antecedentes de SIRI* dataset from the PGN records the **sanctions imposed on convicted public officials** between 2003 and 2019, though the administrative data is much more complete after 2013, so we focus on the 2013-2019 period. These sanctions are the most severe type of penalty imposed, since they do not incorporate the time or hassle associated with *any* investigation. However, they allow us to characterize an upper bound on $P(\theta, r)$ and characterize that bureaucrats do indeed have "skin in the game" when their entities are investigated.

Table A2 reports the rate of investigation of entities led by elected officials in the pre-treatment period and in 2019 is non-trivial. We focus on whether any investigation was conducted and the number of investigations that proceeded to a disciplinary investigation (or further). Since further progress depends on other actors (prosecutors and courts), we emphasize the steps in the process that are controlled by the PGN. Table A2 shows that the vast majority of entities with elected officials were investigated at some point between 2016 to 2019. However, the rate of investigation in any given year is lower, ostensibly due to capacity constraints. There is also significant variation across institutions: departmental *gobernaciones* and municipal *alcaldías* are investigated at a substantially higher rate than their respective legislative counterparts.

		Investigations from 2016-2019		Investigations in 2019	
Entity class	n	Investigated	Disciplinary investigation	Investigated	Disciplinary investigation
Presidency	1	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Congress	2	2 (100%)	1 (50%)	2 (100%)	0 (0%)
Departmental Gobernación	32	29 (90.6%)	3 (9.3%)	29 (90.6%)	2 (6.3%)
Departmental Assembly	32	23 (71.8%)	7 (21.9%)	14 (43.8%)	2 (6.3%)
Municipal Alcaldía	1096	1019 (93.0%)	389 (35.4%)	883 (80.6%)	225 (20.5%)
Municipal Council	1096	455 (41.5%)	140 (12.7%)	252 (23.0%)	61 (5.6%)
Total	2259	1528 (67.6%)	540 (23.9%)	1180 (52.2%)	290 (12.8%)

Table A2: n refers to the number of entities in the experimental (non-pilot) sample. All percentages are expressed as the share of entities in this sample.

Table A3 reports the distribution of sanctions across different types of public sector officials. Note that we do not have denominators for the total number of bureaucrats or police/military members so we cannot calculate the relative likelihood of sanction. Nevertheless, bureaucrats (excluding police/military officials) represent the second largest category of sanctioned individuals. Sanctions typically consist of removal from one's post and some period in which a sanctioned individual cannot serve in the public sector. The median period for which individuals are banned from the public sector is 10 years (and the mean is 11.6 years).

	Number of officials sanctioned	
Official type	2013-2019	2019
Bureaucrat	3,501 (28.8%)	572 (34.3%)
Elected official	370 (3.0%)	64 (3.8%)
Judge	63 (0.5%)	11 (0.7%)
Police/military	8,232 (67.7%)	1,019 (61.2%)

Table A3: Count of sanctioned officials in two time periods, 2013-2019 (when the SIRI data is most complete) and in 2019 (one year before the experiment). Percentages report the share of sanctioned officials in each category.

A2 Theoretical results and examples

A2.1 Proofs

Proof of Remark 1: A bureaucrat reports information $r \in \mathbb{R}$ at cost $\tilde{e} > 0$ if and only if:

$$\begin{aligned} -E[\rho(r \in \mathbb{R})P(r \in \mathbb{R};\theta)] - c(\tilde{e}) &> -E[\rho(\emptyset)P(\emptyset;\theta)] - 0 \\ E[\rho(\emptyset)P(\emptyset;\theta)] - E[\rho(r \in \mathbb{R})P(r \in \mathbb{R};\theta)] &> c(\tilde{e}). \end{aligned}$$

This implies that the bureaucrat reports some information ($r \in \mathbb{R}$) if and only if they perceive that oversight ($E[\rho(r)P(r;\theta)]$) is sufficiently more costly if they do not report ($r = \emptyset$), since $c(\tilde{e}) > 0$. ■

Proof of Remark 2: We examine the bureaucrat's choice of effort, e, by considering two cases. First, suppose that oversight $\rho(r)P(r;\theta)$ is linear in r, such that $\rho(r)P(r;\theta) = \alpha_0 + \alpha_1 r + f(\theta)$. In this case, the bureaucrat's expected utility simplifies to:

$$\begin{aligned} -E[\alpha_0 + \alpha_1 r + f(\theta)] - c(e) &= -E[\alpha_0 + \alpha_1(\theta + d + \varepsilon) + f(\theta)] - c(e) \\ &= -\alpha_0 - \alpha_1(\theta + d) - f(\theta) - c(e), \end{aligned}$$

which follows because $E[\varepsilon] = 0$. Evaluating $\frac{\partial E[\rho(r)P(r;\theta)]}{\partial e} = 0$ and $\frac{\partial c(e)}{\partial e} > 0$ (by assumption) shows that effort (e), enters expected utility only through the cost of effort, $c(e)$, but not through oversight. Second, suppose that oversight $\rho(r)P(r;\theta) = g(r)$ where $g(\cdot)$ is a non-linear function of r. Due to the non-linearity of $g(\cdot)$, we know that $E[g(r)] \neq g(E[r])$. Evaluating $E[g(r)]$ depends on the variance of ε, $\sigma^2(e)$. Thus, when oversight is not linear in r, $\frac{\partial E[\rho(r)P(r;\theta)]}{\partial e}$ is a function of $\sigma^{2'}(e) < 0$, which implies that optimal effort is (generically) a function of expected oversight. As in the first case, $\frac{\partial c(e)}{\partial e} > 0$ (by assumption). Collectively these two cases show that conditional on reporting ($e > 0$), optimal effort is a function of the cost of effort and can be a function of expected oversight, depending on the functional form of $\rho(r)P(r;\theta)$.

Now consider the bureaucrat's choice of d. First, suppose that oversight does not depend on reported scores, such that $\rho(r) = \alpha$, $P(r;\theta) = \beta + h(\theta)$. This implies that:

$$\begin{aligned} -E[\rho(r)P(r;\theta)] &= -E[\alpha(\beta + h(\theta))] \\ &= -\alpha\beta - \alpha h(\theta) \end{aligned}$$

It is clear from this expectation that $\frac{\partial E[\rho(r)P(r;\theta)]}{\partial d} = 0 \forall d$. Given our assumption that the bureaucrat chooses $d = 0$ when indifferent between distorting ($d \neq 0$) and not distorting ($d = 0$) scores, the bureaucrat chooses $d = 0$. Second, suppose that $d^* \neq 0$. In this case, there exists some d_1 for which $\frac{\partial E[U_B(d,e;\theta)]}{\partial d} \neq 0$ and some d_2 for which $\frac{\partial E[U_B(d,e;\theta)]}{\partial d} = 0$. Since $\frac{\partial c(e)}{\partial d} = 0$, it must be the case that $\frac{\partial E[\rho(r)P(r;\theta)]}{\partial d} \neq 0$ which implies that oversight relies on reported scores r (since $r = \theta + d + \varepsilon$). Collectively these cases imply that $d \neq 0$ can only be optimal if and only if oversight depends on r. ■

A2.2 Examples

This section proposes two examples that adopt explicit parameterizations of $\rho(r)$ and $P(r;\theta)$ for the purposes of illustration. In both cases, examples suppose that $\sigma^2(e) = \frac{1}{e}$ and that $c(e) = \frac{e^2}{2}$.

Example 1: Monitoring is uniform in r, but punishment depends on deviations from θ. Suppose that we have $\rho(r) = \alpha \in (0,1)$, $P(r;\theta) = (\theta - r)^2$ for $r \in \mathbb{R}$, $P(\emptyset;\theta) = k > 0$. Here, the bureaucrat's expected utility is:

$$\begin{aligned} E[U_B(d,e;\theta)] &= -E[\alpha(\theta - r)^2] - c(e) \\ &= -E[\alpha(\theta - (\theta + d + \varepsilon))^2] - c(e) \\ &= -\alpha d^2 - \alpha\sigma^2(e) - c(e) \end{aligned}$$

Maximizing with respect to d and e we have:

$$\begin{aligned} \frac{\partial E[U_B(d,e;\theta)]}{\partial d} &= -2\alpha d = 0 \to d^* = 0 \\ \frac{\partial E[U_B(d,e;\theta)]}{\partial e} &= -\frac{\alpha}{e^2} - e = 0 \to -\alpha\sigma^{2'}(e^*) = c'(e^*) \end{aligned}$$

Substituting in $\sigma^{2'}(e)$ and $c'(e)$, we have $\frac{\alpha}{e^2} = e$, which yields $e^* = \alpha^{1/3}$. Thus, the bureaucrat chooses $d = 0, e = \alpha^{1/3}$ if:

$$\begin{aligned} -\frac{\alpha}{\alpha^{2/3}} - \frac{\alpha^{2/3}}{2} &> -\alpha k \\ k &> \alpha^{-2/3} + \frac{\alpha^{-1/3}}{2}, \end{aligned}$$

else, they chose $e = 0$ and do not report.

Example 2: Monitoring and punishment are both linear in r. Suppose that we have $\rho(r) = \alpha_0 + \alpha_1 r$ for $r \in \mathbb{R}$, $P(r;\theta) = \beta_0 + \beta_1(r - \theta)$ for $r \in \mathbb{R}$, $\rho(\emptyset) = j \in (0,1)$, and $P(\emptyset;\theta) = k > 0$.[1] Assume that $\alpha_1 \neq 0$

[1] Assume that α_0 and α_1 are constrained such that $\rho(r) \in (0,1)$.

and $\beta_1 \neq 0$ and that these parameters could be negative or positive. The bureaucrat's expected utility is:

$$E[U_B(d,e)] = -\beta_0\alpha_0 - \alpha_0\beta_1 d + \alpha_1\beta_0 d + \alpha_1\beta_1\sigma^2(e) + \alpha_1\beta_1(\theta+d)^2 - \alpha_1\beta_1\theta(\theta+d) - c(e)$$

Maximizing, we have:

$$\frac{\partial E[U_B(d,e)]}{\partial d} = -\alpha_0\beta_1 + \alpha_1\beta_0 - \alpha_1\beta_1 + 2\alpha_1\beta_1(d+\theta) = 0 \Rightarrow d^* = \frac{-\alpha_1\beta_0 + \alpha_0\beta_1 + \alpha_1\beta_1 - 2\alpha_1\beta_1\theta}{2\alpha_1\beta_1}$$
$$\frac{\partial E[U_B(d,e)]}{\partial e} = \alpha_1\beta_1\sigma^{2'}(e) - c'(e) = 0 \Rightarrow \alpha_1\beta_1\sigma^{2'}(e^*) = c'(e^*)$$

Substituting in $\sigma^{2'}(e)$ and $c'(e)$, we have $-\frac{\alpha_1\beta_1}{e^2} = e$, which yields $e = (-\alpha_1\beta_1)^{1/3}$. Thus, the bureaucrat chooses $d = \frac{-\alpha_1\beta_0+\alpha_0\beta_1+\alpha_1\beta_1-2\alpha_1\beta_1\theta}{2\alpha_1\beta_1}$ and $e = (-\alpha_1\beta_1)^{1/3}$ if:

$$E\left[U_b\left(\frac{-\alpha_1\beta_0 + \alpha_0\beta_1 + \alpha_1\beta_1 - 2\alpha_1\beta_1\theta}{2\alpha_1\beta_1}, (-\alpha_1\beta_1)^{1/3}\right)\right] \geq E[U_b(0,0)]$$
$$\frac{1}{4}\left[-6\alpha_1^{2/3}\beta_1^{2/3} - \frac{\alpha_0^2\beta_1}{\alpha_1} - 2\alpha_0(\beta_0+\beta_1-2\beta_1\theta) - \frac{\alpha_1(\beta_0^2 - 2\beta_0\beta_1 + \beta_1^2 + 4\beta_0\beta_1\theta)}{\beta_1}\right] \geq jk,$$

else, they choose $d = 0$, $e = 0$ and do not report.

A3 Experimental design

A3.1 Research design timeline

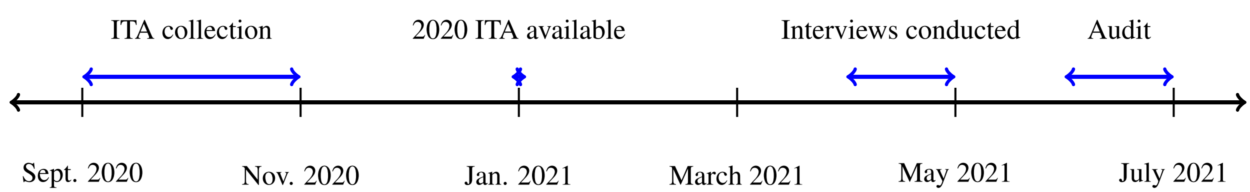

Figure A1: Timeline of research design.

A3.2 Compliance

As a measure of compliance with treatment assignment, we report the proportion of entities for whom the original communication from the PGN was delivered. We can only measure this quantity for entities assigned to direct communication. Table A4 shows that across the full sample, emails were delivered to 94.8% of entities (Column 1). This rate did not vary appreciably across the nudge treatments, as one would expect given that the treatments were revealed within the email. Even among territorial entities, this rate was 94.6%.

A3.3 Inclusion of non-traditional entities

In addition to the near-universe of public sector entities that we describe in the manuscript, the experiment also included a subset of the "non-traditional" entities that are required to fill out ITA under Article 5 of Law

	Direct communication did not bounce					
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.948***	0.995***	0.946***	0.946***	0.996***	0.943***
	(0.003)	(0.005)	(0.003)	(0.007)	(0.004)	(0.007)
Oversight of past completion				-0.001	-0.009	0.000
				(0.006)	(0.009)	(0.006)
Possible future audit				-0.001	-0.010	0.000
				(0.006)	(0.010)	(0.006)
Direct reminder				0.001	0.009	0.001
				(0.006)	(0.009)	(0.006)
Training				0.005	0.009	0.005
				(0.006)	(0.009)	(0.006)
Num. Obs.	5828	210	5270	5828	210	5270
Direct communication	yes	yes	yes	yes	yes	yes
Entities	Public sector	National	Territorial	Public sector	National	Territorial

$^{+}p < 0.1, ^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$

Table A4: Compliance as measured by whether email communication was delivered to entities (did not bounce). Heteroskedasticity-robust standard errors in parentheses.

1712 of 2014 (which we call the "National Transparency Act"). These entities are generally (i) individuals or firms that contract to the state or (ii) political parties and social movements.

Note that in addition to exact blocking on 2019 ITA completion, we also used exact blocking on the traditional/non-traditional distinction. As such, the public-sector entities represented in the main text comprise completely separate blocks from the other entities included in the experiment. Table A5 describes the full experimental sample.

	All Obligated Entities*	**Experimental Entities**	**Audited Entities**
Category	Count (n)	Count (n)	Count (n)
PUBLIC SECTOR	6,556	6,556	2,400
National	237	237	200
Territorial	5,928	5,928	2,200
Undesignated	391	391	0
PRIVATE SECTOR	41,938	5,329	0
PGN Priority	5,329	5,329	0
PGN Non-priority	36,609	0	0
PARTIES/MOVEMENTS	168	168	0
Total	48,662	12,053	2,400

Table A5: Sampling of entities in experiment and audit outcome measurement. *This total omits 62 public sector and 38 private sector entities that were randomly sampled and used in a piloting pre-test of intervention implementation.

A3.4 Covariate balance

Given our multi-arm experimental design, we report two metrics of balance. Three of the four classes of common covariates are generally categorical. For these covariates, we construct indicator variables for common categories (i.e., those with ≥ 36 entities, or two blocks), and regress those category indicators on our main regression specification, reported below:

$$
\begin{aligned}
Y_{ib} = &\beta_0 + \beta_1 \text{Direct Communication}_i + \beta_2 \text{Reminder}_i + \beta_3 \text{Training}_i + \beta_4 \text{Retrospective Oversight}_i + \\
&\beta_5 \text{Prospective Oversight}_i + \epsilon_{ib}
\end{aligned} \tag{1}
$$

From these analyses we report the p-values associated with several test statistics. First, we report p-values from an F-test of the null hypothesis that $\beta_j = 0 \,\forall j \in 1, ..., 5$. This serves as a test of the joint significance of the five treatment indicators. As is suggested by from our results in Figure A2, we fail to reject the null hypothesis (consistent with covariate balance) for all but one covariate examined at the $\alpha = 0.1$ level. Second, upon request, we report the p-values testing the null hypotheses that $\beta_j = 0$ for each $j \in 1, ..., 5$. This tests for imbalance across individual treatment conditions. Again, we do not detect evidence of imbalance in treatment assignment using these tests.

We assess balance on geographic and organizational characteristics as follows:

- **Department**: Each entity is legally registered with a Colombian department. For entities that work in multiple places, this is where the entity is headquartered or encorporated. In these specifications, Y_{ib} indicates whether an entity is based in a given department. We iterate through all of the departments in Colombia.
- **Legal classification of public sector entities (*naturaleza jurídica*)**: For public sector (traditional) entities, this is an indicator for their legal status, i.e., *alcaldía* or local government. We assess covariate balance over common classifications by operationalizing Y_{ib} as an indicator for a given classification. We iterate through all classifications with at least 36 entities. We have preserved the Spanish classification of these entities to reduce confusion in Figure A2.
- **Sector head**: Entities that were in the pure control condition received direction to fill out ITA from sector heads. In the public sector (our main focus), there were only two relevant entity heads, which are exhaustive and mutually exclusive. This implies that balance-related p-values should be identical across the categories. We preserve the Spanish names of these entities to reduce confusion in Figure A2.
- **Past investigation and sanction**: We examine an entity's past history of investigation and disciplinary investigation (only for entities headed by elected politicians) between 2016 and 2019, as well as the history of sanctions imposed against entities/officials within entities between 2003 and 2019. Since the sanction data is more complete after 2013, we also include a 2013-2019 time frame.

A3.5 Intervention materials

We report and translate the content of the direct communication from the PGN to obligated entities assigned to any direct communication treatment in Table A6. All entities received the core information about ITA. The subsequent experimental treatment conditions were randomized (and crossed).

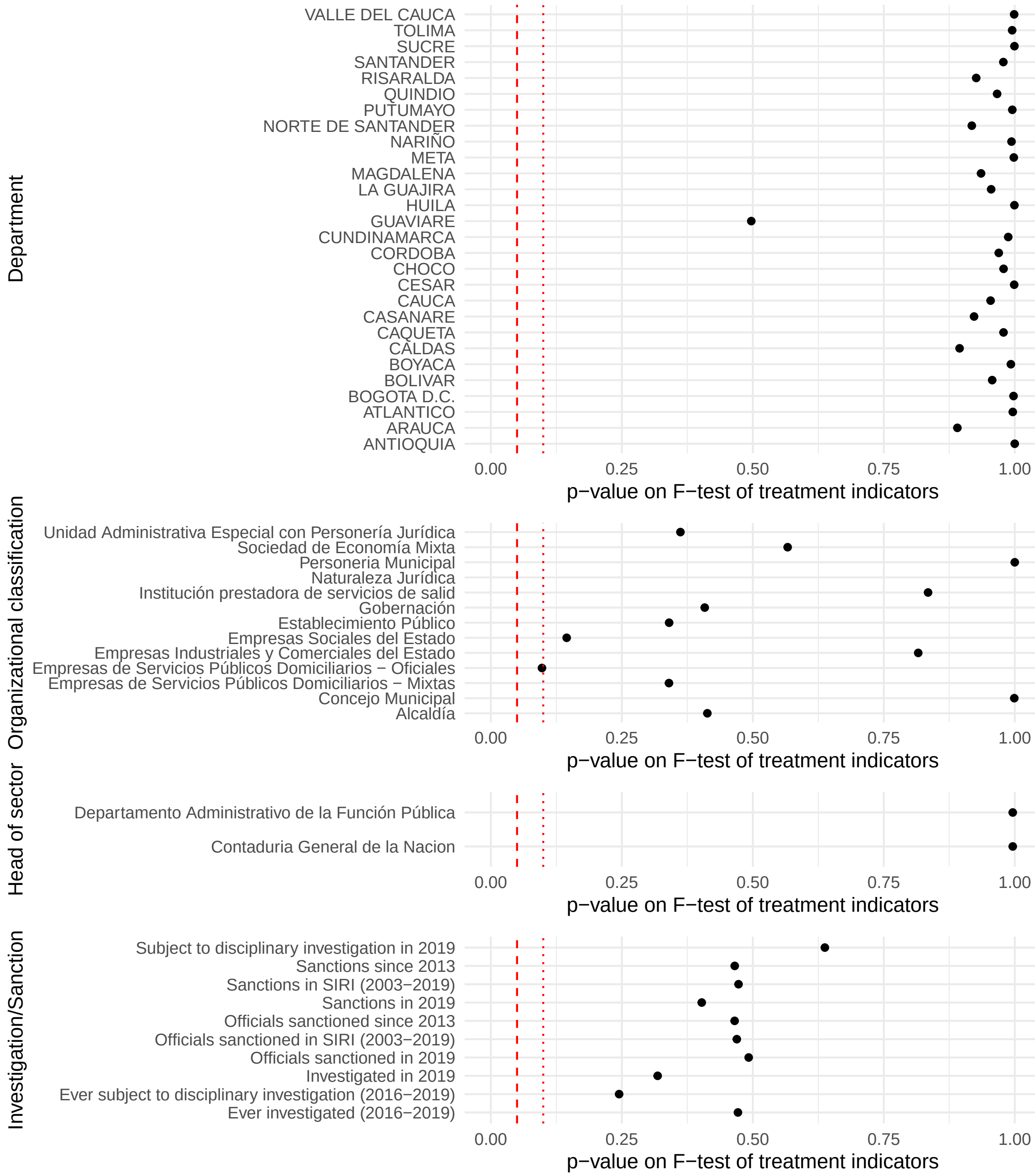

Figure A2: We assess the joint balance of the treatment indicators. This plot graphs the p-values testing the null hypothesis that $\beta_j = 0 \; \forall j \in 1, ..., 5$ in Equation 1 for each outcome indicator along the y-axis. Note that the investigation data includes only public sector entities with elected principals.

Table A6: Original and Spanish translation of the letters sent to entities by treatment status

Treatment condition	**Original message (in Spanish)**	**Translation (English)**
Information about ITA (core)	Como es de su conocimiento, la Procuraduría General de la Nación, de conformidad con lo dispuesto en el artículo 23 de la Ley 1712 de 2014 “Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional”, ha puesto en marcha un sistema de información que permite el registro, seguimiento y monitoreo que automatiza la captura de la información de la “Matriz de cumplimiento normativo de la Ley 1712 de 2014”, a través de un formulario de autodiagnóstico. Dicho sistema es la base del Índice de Transparencia y Acceso a la Información – ITA, cuya primera medición se realizó en 2019. Teniendo en cuenta lo anterior, me dirijo a usted con el propósito de recordarle que la entidad que representa tiene la obligación de diligenciar la Matriz ITA para la medición 2020. La plataforma para diligenciar este formulario estará habilitada entre el 15 de septiembre y el 15 de octubre y puede acceder a ella a través de este enlace: https://apps.procuraduria.gov.co/ita/login/. De igual manera, me permito remitir la Directiva No. 026 del 25 de agosto del año en curso, para su conocimiento, cumplimiento y difusión entre los particulares que estén encargados de realizar esta labor. Para cualquier inquietud o comentarios deberá escribir al correo electrónico: soporteita@procuraduria.gov.co. [Other text] De antemano agradezco mucho la colaboración que usted pueda prestarnos, en orden a promover el cumplimiento de esta Directiva.	As you know, the Office of the Inspector Attorney General, following the provisions of Article 23 of Law 1712 of 2014 "Law of Transparency and the Right of Access to National Public Information,” has implemented a system of information that allows the registration, follow-up, and monitoring that automates the capture of the information of the "Regulatory compliance matrix of Law 1712 of 2014,” through a self-diagnosis form. This system is the basis of the Index of Transparency and Access to Information - ITA, whose first measurement was carried out in 2019. Taking into account the above, I am writing to you to remind you that the entity that you represent must fill out the ITA Matrix for the 2020 measurement. The platform to fill out this form will be enabled between September 15 and 15 October and you can access it through this link: https://apps.procuraduria.gov.co/ita/login/. In the same way, I am sending Directive No. 026 of August 25 of the current year, for your knowledge, compliance, and dissemination among the individuals who are in charge of carrying out this task. For any questions or comments, you should write to the email: soporteita@procuraduria.gov.co. [Other text] Thank you very much in advance for your collaboration in promoting compliance with this Directive.

Table A6 continued from previous page

Training	Adicionalmente, invito a la entidad a visitar la página de la PGN donde se encuentran disponibles dos videos tutoriales donde se explica cómo se debe diligenciar la matriz y la grabación de la capacitación que realizó la PGN en 2019 para instruir y explicar a los sujetos obligados sobre la Matriz ITA. Los videos se encuentran en la parte inferior de la página que se accede a través de este enlace: `https://www.procuraduria.gov.co/portal/ITA.page`.	Additionally, I would like to invite the entity to visit the PGN's website, where two tutorial videos are available explaining how to fill out the matrix and the recording of the training carried out by the PGN in 2019 to instruct and explain to the obligated subjects about the ITA Matrix. The videos are at the bottom of the page, which can be accessed through this link: `https://www.procuraduria.gov.co/portal/ITA.page`.
Oversight (retrospective) for those that complied in 2019	Finalmente, se identificó que la entidad que usted representa diligenció oportunamente la matriz para la medición 2019, dando cumplimiento a la Directiva 006 del 14 de mayo de 2019 proferida por el Procurador General de la Nación. Por este motivo, agradezco su colaboración en la medición anterior y su compromiso para garantizar dentro de la entidad el cumplimiento de lo dispuesto en la Directiva 026 para la medición 2020.	Finally, we identified that the entity you represent filled out the matrix on time for the 2019 measurement, thus complying with Directive 006 of May 14, 2019, issued by the Inspector Attorney General. For this reason, I thank you in advance for your collaboration and commitment to promoting within the entity compliance with the provisions of Directive 026 for 2020 measurement.
Oversight (retrospective) for those that did not comply in 2019	Finalmente, se identificó que la entidad que usted representa no diligenció oportunamente la matriz para la medición 2019, por lo cual se evidencia un incumplimiento a la Directiva 006 del 14 de mayo de 2019 proferida por el Procurador General de la Nación. Por este motivo, agradezco de antemano su colaboración y compromiso para promover dentro de la entidad el cumplimiento con lo dispuesto en la Directiva 026 para la medición 2020.	Finally, we identified that the entity you represent did not fill out the matrix on time for the 2019 measurement, failing to comply with Directive 006 of May 14, 2019, issued by the Inspector Attorney General. For this reason, I thank you in advance for your collaboration and commitment to promoting within the entity compliance with the provisions of Directive 026 for 2020 measurement.
Oversight (prospective)	También quisiera recordarle que el autodiagnóstico y el puntaje que arroja la plataforma no son el resultado definitivo. Dada la importancia del cumplimiento de esta Ley, tal como se realizó con la información del 2019, la información reportada en la medición 2020 pasará por un proceso cuidadoso de revisión de la PGN y puede estar sujeta a un proceso de auditoría de calidad. Lo anterior, de acuerdo con las funciones de vigilancia y prevención de la PGN.	I would also like to remind you that self-diagnosis and scoring that the platform generates are not the final result. Due to the importance of compliance with this Law, as was done with the information reported in 2019, the information reported in the [ITA] 2020 measurement will go through a careful process of oversight by the PGN and may be subject to a quality audit process in accordance with the PGN's functions of surveillance and prevention.

A4 Audited components, microdata

A4.1 Selection into audit sample

Because our independent audit covers only 2,400 of 6,556 public sector entities in the full experimental sample, we consider selection into the audit sample. Per Table 1, it is clear that national entities were sampled at a higher probability than territorial entities. In Tables A7 and A8 we predict selection into the audit sample by estimating the following equation:

$$Y_i = \beta_0 + \boldsymbol{\beta}\mathbf{X}_i + \gamma\text{National}_i + \epsilon_i$$

where $\mathbf{X}_i$ is a matrix of predictors of selection into the audit sample. In Table A7 these predictors include past and current performance on the ITA matrix (in 2019 and 2020, respectively). The inclusion of an indicator for national entities accounts for the differential probability of selection into the audit sample. In Table A8, the predictors include the treatment indicators from the experiment. We do not find evidence of imbalanced selection into the audit. In both tables, all coefficients are very close to zero and we cannot reject the null hypothesis for any predictor.

Figure A3 measures the AMCEs of experimental treatments on the quality measure in the audit (our measure of θ). The outcome ranges from 0-32.4. We find no evidence that the treatments affect audit-measured transparency practices. Our estimates are small in magnitude, and we cannot reject the null hypothesis of no effect for any treatment.

	In audited sample			
	(1)	(2)	(3)	(4)
Submitted data in 2019	-0.001			
	(0.012)			
Entered transparency index system		0.004		
		(0.013)		
Submitted data in 2020			0.008	
			(0.012)	
Transparency index score/100				0.021
				(0.024)
National government entity	0.845***	0.842***	0.840***	0.836***
	(0.025)	(0.024)	(0.024)	(0.026)
Territorial government entity	0.371***	0.370***	0.369***	0.373***
	(0.007)	(0.007)	(0.007)	(0.008)
Intercept	0.000	-0.002	-0.003	-0.014
	(0.002)	(0.006)	(0.005)	(0.016)
Num. Obs.	6556	6556	6556	4446
Adjusted R^2	0.070	0.070	0.070	0.065
Sample	Public sector	Public sector	Public sector	Public sector, completed

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Predicting selection into the audit sample as a function of ITA matrix completion (past and present) and scores. The baseline category for the type of government entity is "unclassified," which comprises $n = 391$ entities. Heteroskedasticity-robust standard errors in parentheses.

	In audited sample			
	(1)	(2)	(3)	(4)
Direct communication	-0.004	-0.004	-0.003	-0.004
	(0.023)	(0.022)	(0.021)	(0.021)
Oversight of past completion	0.019	0.019	0.019	0.019
	(0.013)	(0.012)	(0.012)	(0.012)
Possible future audit	-0.003	-0.004	-0.003	-0.004
	(0.013)	(0.012)	(0.012)	(0.012)
Direct reminder	0.008	0.011	0.008	0.011
	(0.013)	(0.012)	(0.012)	(0.012)
Training	0.010	0.011	0.010	0.011
	(0.013)	(0.012)	(0.012)	(0.012)
National government entity		0.845***		0.893***
		(0.024)		(0.118)
Territorial government entity		0.371***		0.437***
		(0.006)		(0.111)
Num.Obs.	6556	6556	6556	6556
Adjusted R^2	0.000	0.058	0.070	0.074
Sample	Public Sector	Public Sector	Public Sector	Public Sector
Block FE			yes	yes

* p < 0.1, ** p < 0.05, *** p < 0.01

Table A8: Predicting selection into the audit sample as a function of experimental treatments. The baseline category for the type of government entity is "unclassified," which comprises $n = 391$ entities. Heteroskedasticity-robust standard errors in parentheses.

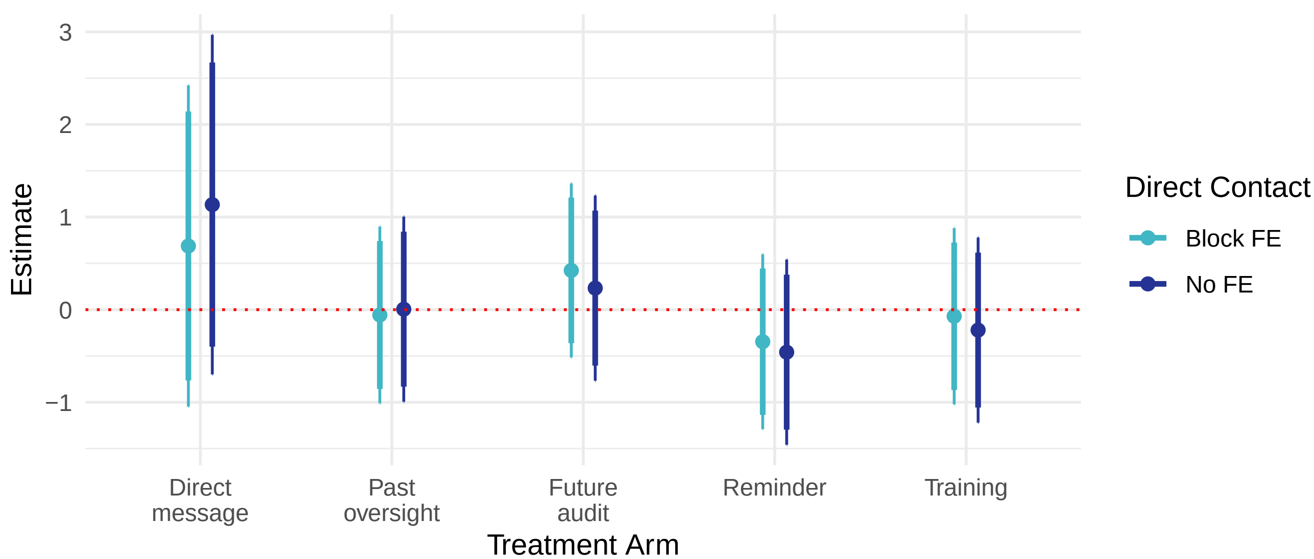

Figure A3: Estimates of the AMCE of experimental treatments on audit-measured quality within the subset of audited entities.

A4.2 Audited components

To keep the audit manageable, we defined a list of ten components (each consisting of at least one item) to be audited by the independent firm. These were selected based on two criteria:

1. *Relevance*: Components which contain information that citizens would be most likely to consult or use.

2. *Feasibility*: Components that are comparable across entity classes and could be systematized by auditors.

The first criterion led to a list of 6 components. For the latter criterion, we mainly relied on the results of a pilot that we conducted with the goal of assessing the feasibility of the audit itself, and the validity of the instrument. After filtering out those components in which responses were not comparable across entities or were excessively time or data consuming (i.e., components that required downloading and opening multiple files for validation), we defined the following list of components to audit. Note that all components come from the original ITA matrix form.

1. *Transparency button*: There is a designated "Transparency" section on the home page of the obliged subject's website.
2. *Mechanisms for citizen service*: The entity must publish a section where the information on the mechanisms through which the citizen can contact the entity is described.
3. *Form for receiving citizen information*: This component refers to the characteristics of the form for receiving requests for public information. It requires to have the following fields: (a) Applicant type (person or company); (b) First name; (c) Surname; (d) ID Type; (e) Identification number; (f) Company's name; (g) Tax ID number (NIT); (h) Country; (i) Department; (j) Municipality; (k) Address; (l) Email; (m) Landline and / or mobile phone; (n) Content of the request; (o) Files or document attachments; (p) Option to choose the response channel; and (q) Information on possible costs associated with the response.
4. *Open data*: The entity must publish the data generated by the entity on its website. They must at least have the following: (a) Index of reserved and classified public information; and (b) Records of Information Assets. Both must be published in open data. NOTE: The publication of these data, regardless of the file format in which it is found (Word, Excel, CSV), must be available in an accessible and reusable way.
5. *Frequently asked questions*: The entity must publish a list of frequently asked questions with the respective answers, related to the entity, its management and the services and procedures it provides.
6. *Mission and vision*: The entity must publish information on its mission and vision in accordance with the creation or restructuring standard or as defined in the entity's quality management system.
7. *Assigned general budget*: The entity must publish the general budget assigned for each fiscal year. (Many times the entities publish the decree defining that budget.)
8. *Management, evaluation, and audit reports*: The entity must publish the management, evaluation and audit reports, including the budget year.
9. *Publication of audit information*: The entities that contract with public resources, or public and private resources, must publish in the SECOP the information of their contractual management with charge of public resources.
10. *Means of tracking public information requests*: Entities must provide a tracking number or code for so that petitioners can follow up on public information requests.

While the components we select are of high relevance to citizens, one might ask if they are similarly of high relevance or importance to the central government. We code components in terms of their relevance to the central government. We assume that all components have *some* value since the central government (PGN) chooses to include them. Since the central government determines the weighting scheme, we rely on those weights to infer component importance. Specifically, we define importance to the central government as "high" if the weight assigned to an component is above the median, and "low" if it is below the median across all components.[2]

Table A9 reports the distribution of ITA components by their importance to citizens and the central government. The table shows that our audit focuses on components that are important to *both* citizens and the central government. Of the 32.4 points of the ITA index that we audit, 30.2 points (93.2%) are classified as of high importance to the central government. Note that this classification is also reflected in the implicit weighting of the index. While we audit just 10 of 59 components (16.9%), these components represent 32.4% of the overall index.

		Total		Audited		Not audited	
Central government	**Citizen**	**Components**	**Weight**	**Components**	**Weight**	**Components**	**Weight**
High	High	11	46.60	5	27.20	6	19.40
	Low	21	38.20	1	3.00	20	35.20
Low	High	9	4.60	4	2.20	5	2.40
	Low	18	10.60	0	0.00	18	10.60
	Total	59	100	10	32.40	49	67.60

Table A9: Distribution of components by relevance to central government and citizens. The audit components were selected on the basis of importance to citizens. These components are also disproportionately important to the central government in the context of intergovernmental relationships with decentralized entities.

In comparing component-level responses to their audit results to measure discrepancies between reported and true scores, we rely on ITA responses present in the public microdata that records responses to each component. When we compare entities for which the PGN has recorded a score to those in the public microdata, we observe some discrepancies. Specifically, there are fewer entities in the public microdata than entities that completed the ITA matrix according to the PGN. This is evident in the lower left cell of Table A10, where nearly 20% of audited entities completed the matrix but are not present in the microdata. A further 3.7% of the sample did not complete the ITA matrix per PGN's match but is in the public microdata. Because bureaucrats self-reported entity names, which often do not match the administrative records, the PGN and the research team conducted separate hand matches between the data inputs and the scores. These 3.7% of entities is suggestive of the lack of overlap in these matches. Ultimately, this suggests that measurement error due to misattribution of scores to entities is quite limited. Our primary concern, which we discuss at greater length when interpreting results, is the absence of some entities from the public microdata.

[2]Because the final weighted score is derived from weights at three levels—the category, subcategory, and component level—aggregation at the component level produces a slight imbalance in the number of components per category.

	Not in public microdata	In public microdata
Did not complete ITA (PGN measure)	615 (25.6%)	89 (3.7%)
Completed ITA (PGN measure)	478 (19.9%)	1,218 (50.8%)

Table A10: Confusion matrix for PGN data versus public microdata.

A5 Semi-structured interviews

We conducted interviews with officials in different public sector entities that filled out the ITA in late 2020. We conducted these interviews in 2021 after the microdata became available. We identified the official responsible for submitting an entity's data available from this microdata. We sent invitations to participate in semi-structured interviews about ITA and data reporting in general. The response rate was 7%.

Our sampling strategy was as follows. First, we identified contacts at public sector ("traditional") entities from public microdata available from the PGN. We then eliminated any contacts within the PGN. These were typically individuals that assisted with submission of the ITA upon request by officials in a given entity. To sample, we stratified along three dimensions: (1) assignment to control versus direct message treatment assignment; (2) reported score within three bins: $\{\leq 20, \in (20, 80), \geq 80\}$; and (3) whether the entity has an elected principal. Entities are classified as having an elected principal if the principal was selected in an election, which includes: *alcaldías* (local governments); *consejos* (local councils); *gobernaciones* (department governments); *asambleas* (department councils); the presidency; and both houses of Congress (*Cámara de Representantes* and *Senado*). This stratification helps to ensure that we are considering a variety of entities in the qualitative analysis.

One limitation of the survey evidence is that we do not observe officials who *declined* to submit ITA. Nevertheless interviews with officials that chose to submit data allows us to study how they understand oversight and the role of the PGN.

A6 Ancillary experimental analyses

We report several ancillary analyes frrom the experiment, as follows:

1. Figure A4 plots the conditional AMCEs on completion of the 2020 ITA index (left) and the difference in these conditional AMCEs by sector of entities (right). Public sector entities include PGN-designated "traditional" public sector entities. Other entities include private firms, political parties, and social movements. The left plot reports the conditional AMCEs. The estimates in the left panel come from Figure 3. We estimate the difference in conditional AMCEs using the estimators γ_j in the following OLS specification:

$$Y_{ib} = \sum_j \beta_j Z_{ij} + \sum_j \gamma_j \text{Public Sector}_i Z_{ij} + \boldsymbol{\psi}_b + \epsilon_{ib} \tag{2}$$

   Note that j indexes the treatment arms of the factorial design and Z_{ij} is an indicator variable capturing assignment to treatment j. The right panel of Figure 3 plots our estimates of the estimated γ_j's.

2. Figure A5 reports conditional AMCEs on completion of the ITA index among public sector entities, as a function of 2019 ITA matrix completion (i.e., one the lagged dependent variable). We calculate

estimated conditional AMCEs and differences in CAMCES from an estimator analogous to (2) where the moderator is 2019 index completion.

3. Figure A6 reports the association between each of the treatment conditions and the reported scores, conditional on completion of ITA. The left-hand side plot reports the estimates from Panel B of Table 3. The right-hand side plot reports differences in these estimates as estimated through an estimator analogous to (2).

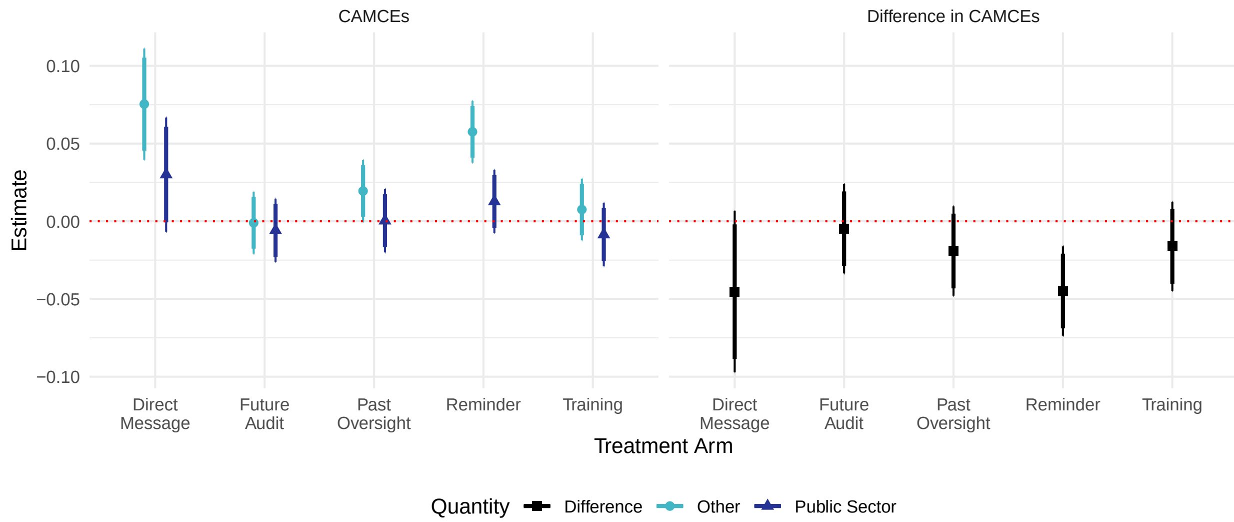

Figure A4: Conditional AMCEs on transparency index completion, among public sector and other (non-public sector) entities (left panel). Differences in conditional AMCEs between public sector and other entities (right). 90% (thin) and 95% (thick) confidence intervals, based on heteroskedasticity-robust standard errors.

A7 Evaluating monotonicity of selection into reporting

In Table 3, we show that direct communication from the PGN increased rates of reporting by entities assigned to treatment. In the decomposition of the post-treatment estimand on scores in Appendix A8, we invoke Lee (2009) bounds to estimate interval estimates of the treatment effects on always-reporters. Lee bounds assume monotonicity (or no defiers) on selection into reporting. In this analysis, we provide support for that assumption. To evaluate the assumption of monotonicity, we use machine learning to estimate CATEs across a large set of pre-treatment covariates. Specifically, we use a generalized random forest model proposed by Athey, Tibshirani, and Wager (2019). We employ each of the following covariates (as binary indicators for each level): department, level of entity (national or territorial), administrative classfication (per Colombian government classification scheme), central/decentralized administration (per Colombian government classification scheme), ITA completion in 2019, type of entity (for large categories), and an indicator for an organization with a tax identification number (in DIAN). This yields a matrix of 61 predictors.

Figure A7 depicts the predicted CATEs for each entity. It shows that while 27% of the estimates are negative, none are statistically distinguishable from zero at the $\alpha = 0.05$ level. In contrast, we estimate precisely-estimated positive treatment effects for 13.5% of entities. We interpret our inability to detect (statistically) a

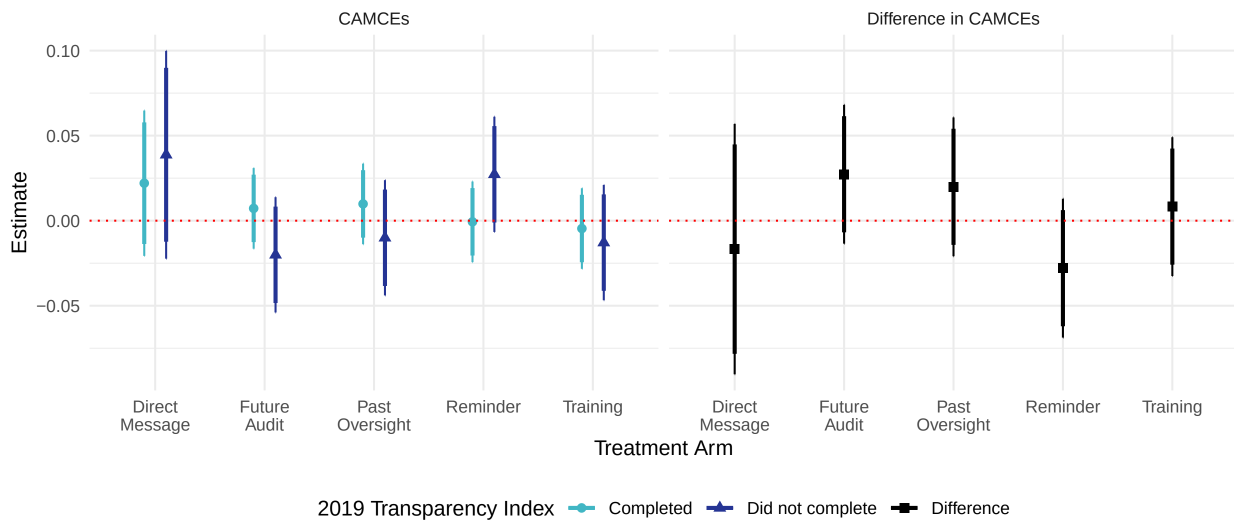

Figure A5: Conditional AMCEs on transparency index completion, among public sector entities a function of index completion in 2019. Differences in conditional AMCEs for 2019 reporters and non-reporters (right). 90% (thin) and 95% (thick) confidence intervals,based on heteroskedasticity-robust standard errors.

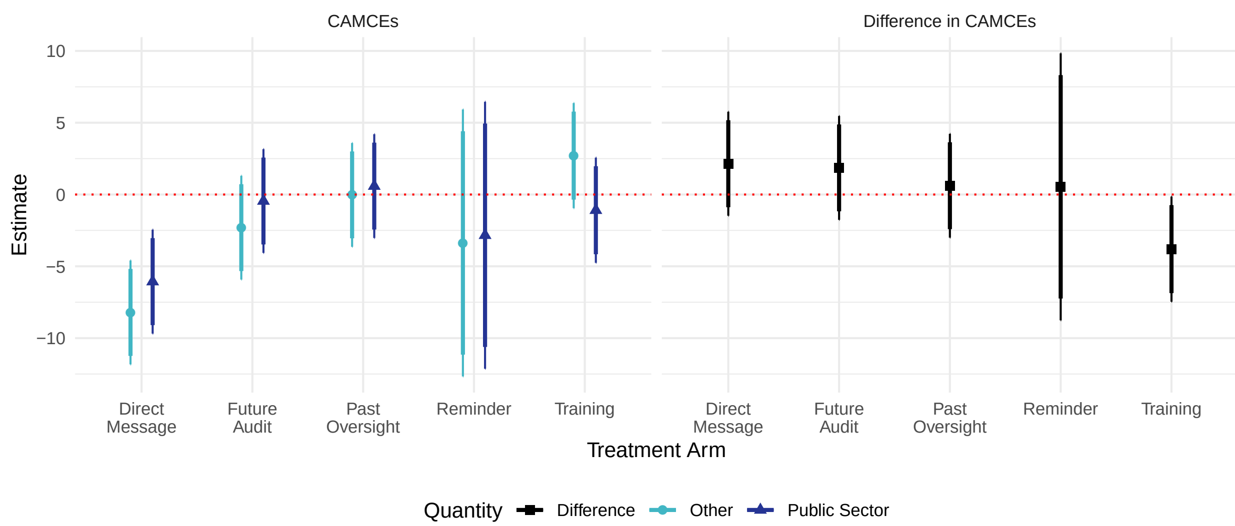

Figure A6: Conditional association between treatments and transparency index scores, among public sector and other (non-public sector) entities (left). Differences in conditional associations between public sector and other entities (right). 90% (thin) and 95% (thick) confidence intervals,based on heteroskedasticity-robust standard errors.

negative treatment effect for any institution as evidence consistent with our assumption of monotonicity.

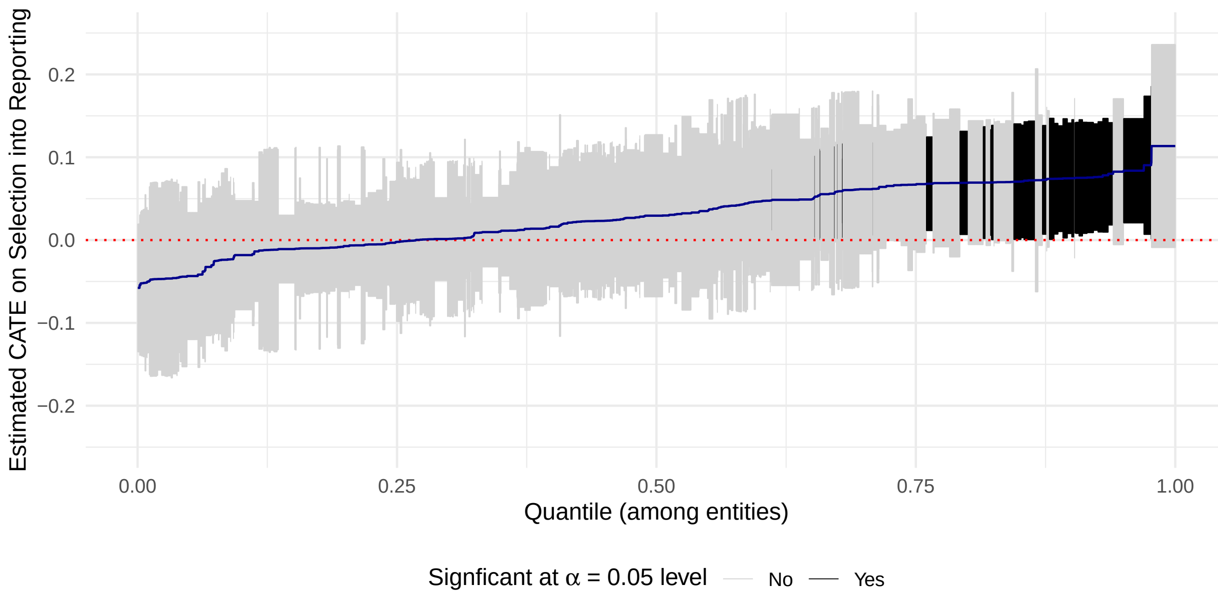

Figure A7: Predicted CATE for each entity ($n = 6,556$) estimated by generalized random forest estimator. Each interval represents a 95% confidence interval. The navy line is the ECDF of the CATEs.

A8 Decomposition of the post-treatment estimand

Denote the outcome, reporting $R_i \in \{0, 1\}$ and the reported score $S_i \in [0, 100]$. Table 3 reports estimates of the post treatment estimand:

$$\mathcal{P} \equiv E[S(Z = 1)|R(Z = 1) = 1] - E[S(Z = 0)|R(Z = 0) = 1]$$

Consider a binary treatment $Z \in \{0, 1\}$ where Z corresponds to direct contact from the PGN. We invoke an assumption of monotonicity: this implies that no entity that would have reported without direct communication failed to report because of the direct communication. Under the assumption of monotonicity, there are three potential (causal) types in the data: always reporters (indicated by $j = A$), never reporters (indexed by $j = N$), and if-treated reporters (indexed by $j = T$). We denote the shares of each type as π_j where $\sum_j \pi_j = 1$.

Now, consider each term in $E[S(Z = 1)|R(Z = 1) = 1] - E[S(Z = 0)|R(Z = 0) = 1]$:

$$E[S(Z = 1)|R(Z = 1) = 1] = \frac{\pi_A}{\pi_A + \pi_T} E[S(Z = 1)|j = A]| + \frac{\pi_T}{\pi_A + \pi_T} E[S(Z = 1)|j = T] \tag{3}$$

$$E[S(Z = 0)|R(Z = 0) = 1] = E[S(Z = 0)|j = A] \tag{4}$$

Simplifying terms we can express $\mathcal{P}$ as:

$$\begin{aligned}\mathcal{P} = & \underbrace{\frac{\pi_A}{\pi_A + \pi_T} \left(E[S(Z=1)|j=A] - E[S(Z=0)|j=A]\right)}_{\text{Change in scores reported}} + \\ & \underbrace{\frac{\pi_T}{\pi_A + \pi_T} \left(E[S(Z=1)|j=T] - E[S(Z=0)|j=A]\right)}_{\text{Change in composition of reporters}} \\ = & \frac{\pi_A}{\pi_A + \pi_T} CATE + \frac{\pi_T}{\pi_A + \pi_T} \left(E[S(Z=1)|j=T] - E[S(Z=0)|j=A]\right)\end{aligned} \tag{5}$$

where $CATE$ is the conditional average treatment effect among always-reporters ($j = A$). We use Lee (2009) trimming bounds to bound $CATE \in [CATE_L, CATE_U]$. With these bounds, Equation (5) similarly provides bounds on the average score of if-treated reporters, $E[S(Z = 1)|T]$. Rearranging Equation (5) we derive the bounds:

$$\begin{aligned}E[S(Z=1)|T] = \Big[& \frac{\mathcal{P} - \frac{\pi_A}{\pi_A+\pi_T} CATE_U + \frac{\pi_T}{\pi_A+\pi_T} E[S(Z=0)|j=A]}{\frac{\pi_T}{\pi_A+\pi_T}}, \\ & \frac{\mathcal{P} - \frac{\pi_A}{\pi_A+\pi_T} CATE_L + \frac{\pi_T}{\pi_A+\pi_T} E[S(Z=0)|j=A]}{\frac{\pi_T}{\pi_A+\pi_T}} \Big]\end{aligned} \tag{6}$$

Note that it is straightforward to calculate point estimates of $\mathcal{P}$, π_A, π_T, and $E[S(Z = 0|j = A)]$. With these point estimates and the Lee (2009) bounds, $CATE_L$ and $CATE_U$, it is straightforward to bound $E[S(Z = 1)|T]$.

Table A11 reports our interval estimates of the CATE of treatment on the scores reported by always-reporters. We generate uncertainty estimates using bootstrapping. Table A11 reports the 95$^{\text{th}}$ percentile of the upper bound estimates, $(Q_{95}(\widehat{CATE}_U)$. Because this decomposition employs ATEs instead of AMCEs in this analysis, we consider multiple definitions of treatment. In Panel A we consider the ATE of direct communication treatment (pooled over content) versus pure control, as in the top interval estimates in Figure 2. In Panel B, we consider treatment as direct communication and any individual manipulation of content versus pure control, as in the subsequent four estimates in Figure 2. Note that the number of observations in the treatment group decreases substantially from Panel A to Panel B. Because we want to test whether the upper bound of the interval estimate is negative, we assess the sign of the upper bound of the 95% confidence interval on the upper bound of our interval estimate.

A9 Audit data: ancillary analyses

We conduct several ancillary analyses of the audit data, as described below:

1. We conduct a simulation to understand the properties of non-systematic error given the distribution of audit scores, θ. In the simulation, we assume that all error is non-systematic and that all entities have the same probability of error on a given ITA matrix item. We take the audit-measured compliance and vary the error rate from 0 to 0.5 and randomly simulate 5,000 iterations of the "observed" data under a given error rate. We then calculate (i) the mean score, which allows us to assess the direction of bias;

Direct communication content				Lee Bounds		Upper bound
Past	Future	Reminder	Training	$\widehat{CATE}_L$	$\widehat{CATE}_U$	$Q_{95}(\widehat{CATE}_U)$
PANEL A: DIRECT COMMUNICATION (ANY)						
Any	Any	Any	Any	-7.97	-5.23	-1.37*
PANEL B: DIRECT COMMUNICATION ALONG EACH MARGIN						
✓	Any	Any	Any	-7.67	-4.59	-0.58*
Any	✓	Any	Any	-8.19	-5.65	-1.39*
Any	Any	✓	Any	-9.38	-5.60	-1.69*
Any	Any	Any	✓	-8.52	-6.14	-1.81*

Table A11: Lee trimming bounds on the conditional average treatment effect (CATE) among the principal stratum of always-reporter entities. $\widehat{CATE}_L$ and $\widehat{CATE}_U$ correspond to the lower and upper bounds, respectively. To rule out compositional effects in isolation, we test the one-tailed null hypothesis, $H_0 : \widehat{CATE}_U \geq 0$. The 95th percentile of the bootstrapped distribution of $\widehat{CATE}_U$, $(Q_{95}(\widehat{CATE}_U))$ allows us to test this hypothesis. * corresponds a rejection of H_0 at the $\alpha = 0.05$ level.

(ii) the conditional standard deviation in scores, which we compare to Figure 5; and (iii) the predicted distortion, which we compare to Figure 4. These results are in Figure A8. The simulation shows that non-systematic error biases scores down, and that the patterns of conditional variance and distortions are inconsistent with a constant level of non-systematic error for all θ.

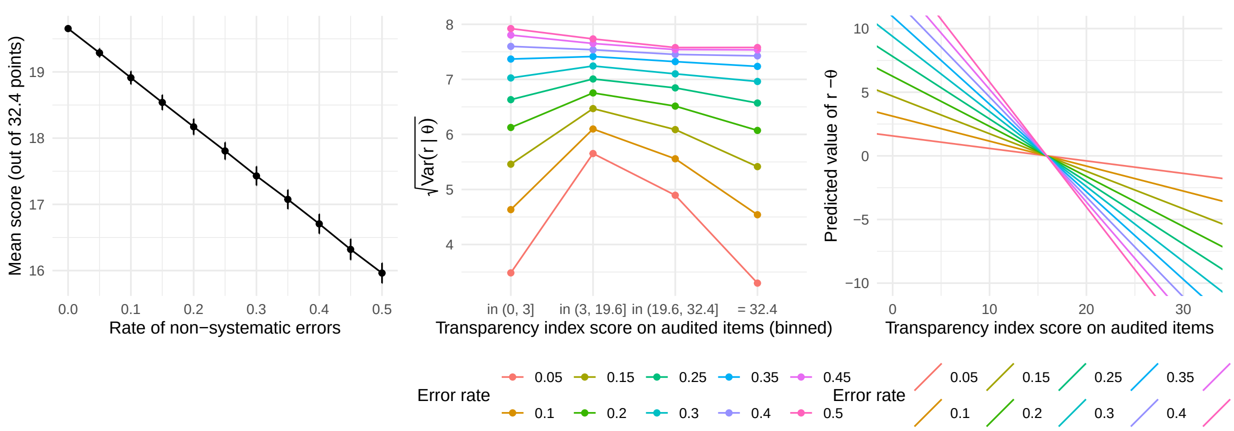

Figure A8: Results from our simulation of non-systematic error with different rates of non-systematic error.

2. We examine the relationship between two measures of administrative capacity and transparency practices and reporting behavior, to assess how administrative capacity influences reporting dynamics, as follows:

   - **Municipal-level capacity**: We use the National Planning Department;s (DNP) index of municipal performance to measure geographical variation in administrative capacity. This measure evaluates local government outputs, which cover a subset of the entities we study. We have this measure for 952 of the 953 municipalities represented in the audit, minimizing missingness.
   - **Entity-level capacity:** We use data from the Administrative Department of Public Administra-

tion's (DAFP) index of institutional performance to assess administrative capacity at the entity level. This index is constructed from self-reports (like the ITA) and is not required for all entities mandated to report ITA data, resulting in higher missingness. We have scores for only 1,210 of 2,400 audited entities.

Table A12 reports the associations between these measures of municipal and institutional capacity and our measures of reporting behavior from the audit. In Columns 1-2, we show that the probability of completion of the ITA matrix increases in both measures of administrative capacity. Columns 3-4 show that audit-measured transparency scores are increasing in municipal capacity. Columns 5-8 report distortions in reporting among the entities that reported ITA transparency scores. First, note that Columns 5-6 document an *increase* in the degree of over-reporting as a function of administrative capacity. Columns 7-8 show that despite this systematic over-reporting by higher-capacity entities, higher-capacity entities also report with *less noise* (greater accuracy) than low-capacity entities.

	Completed ITA $\mathbb{I}(r \neq \emptyset)$		Audit score (θ)		Distortion $(r - \theta)$		| Distortion | $(\mid r - \theta \mid)$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Muncipal capacity	0.018^{+}		0.878^{**}		1.225^{***}		-0.835^{***}	
	(0.009)		(0.268)		(0.182)		(0.163)	
Institutional capacity		0.116^{***}		2.552^{***}		1.664^{***}		-1.502^{***}
		(0.011)		(0.337)		(0.281)		(0.244)
Num. Obs.	2396	1190	2396	1190	1693	960	1693	960
DV mean	0.707	0.807	19.649	24.774	-1.175	-0.676	4.549	4.288
DV std. dev.	(0.455)	(0.395)	(13.864)	(11.073)	(8.883)	(8.136)	(7.719)	(6.946)
DV range	{0, 1}	{0, 1}	[0, 32.4]	[0, 32.4]	[-32.4, 32.4]	[-32.4, 32.4]	[0, 32.4]	[0, 32.4]

$^{+}p < 0.1, ^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$

Table A12: Associations between each of the capacity measures and measured parameters capturing reporting behavior. All regressions are estimated by OLS with heteroskedasticity-robust standard errors.

3. Table A13 considers the possibility of interactions between audit-measured scores on the transparency index and administrative data in determining reporting behavior. It estimates specifications of the form:

$$Y_i = \beta_0 + \beta_1 \text{Audit Score}_i + \beta_2 \text{Capacity}_i + \beta_3 \text{Audit Score}_i \times \text{Capacity}_i + \epsilon_i$$

   We report estimates of β_1, β_2, and β_3 in the table for each outcome (completion, distortions, and absolute value of distortions) and both capacity measures. The principal finding of this analysis is that at low levels of transparency (when the Audit Score is low, or near zero), municipal capacity increases noise (Columns 5-6). However, as the underlying level of transparency increases, noise in reporting decreases in municipal capacity.

4. Figure A9 decomposes the conditional standard deviation of reported scores by audit measured scores (x-axis) and the two measures administrative capacity. Consistent with the idea that low-capacity entities face higher costs of effort, which would lead them to exert less effort, we observe higher-variance in the reported scores from these entities for all subgroups. However, even within the high- and low-capacity medians, we observe similar patterns in the conditional variance. Lower-transparency entities report scores with higher variance.

	Completed ITA $\mathbb{I}(r \neq \emptyset)$		Distortion $(r - \theta)$		$\mid$ Distortion $\mid$ $(\mid r - \theta \mid)$	
	(1)	(2)	(3)	(4)	(5)	(6)
Audit score (0-1 scale)	0.371***	0.176***	-8.575***	-11.066***	-0.283	-3.767***
	(0.021)	(0.038)	(0.545)	(1.002)	(0.531)	(0.968)
Muncipal capacity (standardized)	-0.070***		1.549***		1.739***	
	(0.018)		(0.443)		(0.422)	
Institutional capacity (standardized)		0.141***		1.376+		1.733*
		(0.027)		(0.811)		(0.789)
Audit score × Municipal capacity	0.125***		0.061		-3.475***	
	(0.023)		(0.520)		(0.510)	
Audit score × Institutional capacity		-0.058+		1.058		-4.027***
		(0.032)		(0.909)		(0.891)
Num. Obs.	2396	1190	1693	960	1693	960
DV mean	0.707	0.807	-1.175	-0.676	4.549	4.288
DV std. dev.	0.455	0.395	8.883	8.136	7.719	6.946
DV range	{0, 1}	{0, 1}	[-32.4, 32.4]	[-32.4, 32.4]	[0, 32.4]	[0, 32.4]

$^{+}p < 0.1, ^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$

Table A13: This table investigates interactions between audit-measured transparency scores (rescaled to a 0-1 interval) and both measures of municipal capacity. All regressions are estimated by OLS with heteroskedasticity-robust standard errors.

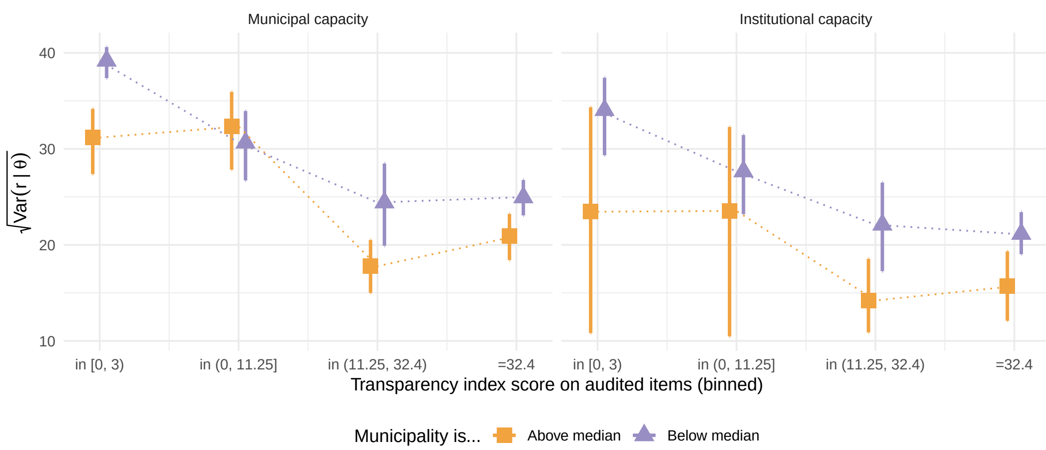

Figure A9: Disaggregation of Figure 5 by medians of the municipal and institutional capacity measures. This figure decomposes the conditional standard deviation in reports by entities above and below the median of state capacity (right).

5. We estimate the conditional average treatment effects of the direct communication treatment at different levels of audit-measured transparency. Here, we estimate CATEs in four bins of audit scores, j. In (7), β_j serve as the estimators of the CATEs in each bin. We examine whether an entity reports and

the distortion in reports (conditional on reporting) as the outcomes, Y_i.

$$\begin{aligned} Y_i = &\sum_{j=1}^{4} \beta_j \text{Direct Communication}_i \times \text{Audit score in } j_i + \\ &\sum_{j=1}^{4} \gamma_j \text{Audit score in } j_i + \kappa \text{National entity}_i + \epsilon_i \end{aligned} \tag{7}$$

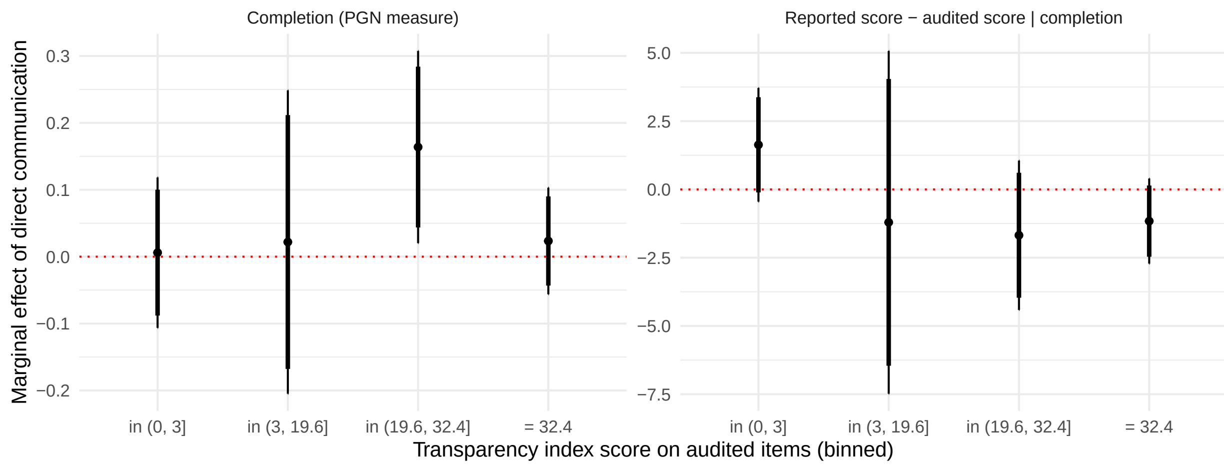

Figure A10: Estimates CATE of direction communication treatment by latent levels of transparency across audited entities in experimental sample. 95% confidence intervals are reported. Note that the right panel conditions the sample on submission, so care should be exercised in interpretation of these estimates as CATEs.

6. We examine the robustness of our findings to the selection of the 10 components for the audit through a series of leave-one-out analyses. Specifically, we estimate the following OLS specifications:

$$\text{Complete}_i = \beta_1^j \text{Audit}_i^{LOO_j} + \beta_2 \text{National}_i + \varepsilon_i \tag{8}$$

$$\text{Reported score}_i^{LOO_j} = \beta_1^j \text{Audit}_i^{LOO_j} + \beta_2 \text{National}_i + \varepsilon_i \tag{9}$$

$$|\text{Distortion}|_i^{LOO_j} = \beta_1^j \text{Audit}_i^{LOO_j} + \beta_2 \text{National}_i + \varepsilon_i \tag{10}$$

Here, $j \in \{1, 10\}$ indexes the left out component. For all leave-one-out quantities with the LOO_j superscript, we rescale the quantity by dividing by its maximum. This maintains comparability across components that are afforded different weight in the ITA index. Note that (8) corresponds to Figure 3; (9) corresponds to Table 4, and (10) roughly corresponds to Figure 5. In Figure A11 we plot estimates of β_1^j from regression specifications (8)-(10). This shows that our estimate and inferences are largely robust to the variation in the choice of index components. The LOO estimates for one component are notable: the exclusion of the component corresponding to the form for receiving citizen information requests (Component #3). Notably, this component is more heavily weighted in the index and is more error-prone for all levels of actual performance, which might be expected since is originally composed of over 20 form items (more than the other components).

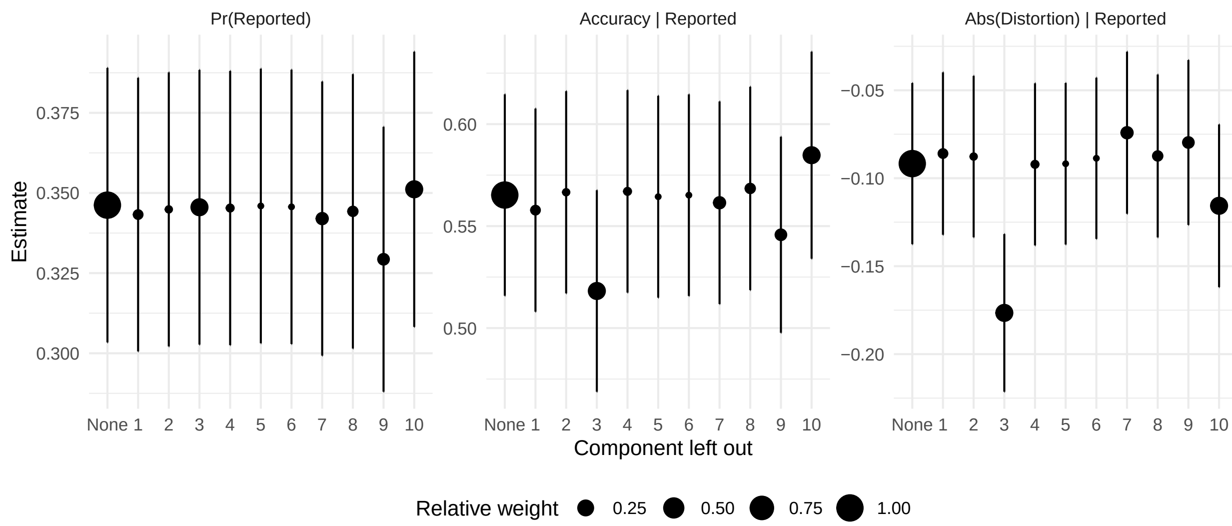

Figure A11: Estimates of β_1^j—the association between our audit measure of θ and the relevant outcome—and the ITA index quantities (for comparison) from equations (8)-(10). The relative weight is the weight attached to the left-out component divided by the total weight of the ten components.

A10 Simulating government use of data

In Figure 7, we consider what the PGN might see under different audit-targeting policies, $\rho(r)$. In Table A14, we report the functional forms that we use for each form of targeting. Note that what varies in the analysis is how the $\rho(r)$ varies in r, not the absolute level or frequency of targeting. For this reason, we represent each non-zero rate of targeting scaled by some arbitrary constant, $k \in (0, 1]$.

Targeting not based on reported scores		**Targeting based on reported scores**	
Targeting strategy	Algorithm	Targeting strategy	Algorithm
All entities	$\rho(r) = k$	Target zero scores	$\rho(r) = \begin{cases} k & \text{if } r = 0 \\ 0 & \text{else} \end{cases}$
Non-respondents only	$\rho(r) = \begin{cases} k & \text{if } r = \emptyset \\ 0 & \text{else} \end{cases}$	Target low scores	$\rho(r) = \frac{k(100-r)}{100}$
Respondents only	$\rho(r) = \begin{cases} k & \text{if } r \in \mathbb{R} \\ 0 & \text{else} \end{cases}$	Target middling scores	$\rho(r) = k\left(\frac{50-\|r-50\|}{100}\right)$
		Target high scores	$\rho(r) = \frac{rk}{100}$
		Target perfect scores	$\rho(r) = \begin{cases} k & \text{if } r = 100 \\ 0 & \text{else} \end{cases}$

Table A14: Algorithms used for simulating government use of data.

Supplementary Appendix: References

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. "Generalized Random Forests." *The Annals of Statistics* 47 (2): 1148–1178.

Hooghe, Elisabeth, Gary Marks, Arjan H Schakel, Sara Niedzwiecki, Sandra Chapman-Osterkatz, and Sarah Shair-Rosenfield. 2021. "Regional authority index (RAI) v. 3.".

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *The Review of Economic Studies* 76: 1071–1102.