The Organizational Dynamics of Administrative Data Production
Appendix

Tara Slough*

July 23, 2026

Contents

*Associate Professor, New York University, tara.slough@nyu.edu

A1 Theoretical Model and Results

A1.1 Model

This model formalizes the verbal intuitions and theoretical results in the main text. The model builds closely upon Crawford and Sobel (1982). Consider the behavior of two bureaucrats, $i \in \{a, b\}$, working in the same organization. Bureaucrat a is tasked with making a report to a non-strategic principal. The report seeks to measure an unobserved state of the world, θ. θ is drawn from a uniform prior with support on the $[0, 1]$ interval.

While bureaucrat a does not observe θ, bureaucrat b does. Bureaucrat a can choose to collaborate, $c \in \{0, 1\}$, which takes the form of seeking a report from b. Collaboration imposes costs on bureaucrat a, $\kappa > 0$, in terms of time, preparation, and/or expectations of future reciprocity by b.

When issuing a report, z_b, the bureaucrat b values both integrity and advocacy. Integrity motives drive the bureaucrat (a or b) to utilize the final report (z_a) to match the state, θ. However b has an advocacy incentive, such that she would prefer to match z_a to her ideal point, $\theta + \beta$, where $\beta \in \mathbb{R}_+$ (without loss of generality) and is common knowledge. The utility of bureaucrat a is given by:

$$u_a(z_a, c; \theta, z_b) = -(\theta - z_a(z_b))^2 - \kappa c,$$

and the utility of bureaucrat b is:

$$u_b(z_b; \theta) = -(\theta + \beta - z_a(z_b))^2.$$

Data collection proceeds according to the following sequence:

1. Bureaucrat a determines whether to seek input, $c \in \{0, 1\}$ from bureaucrat b.
2. If a sought information ($c = 1$), b reports $z_b \in [0, 1]$, else she issues no report ($z_b = \emptyset$).
3. a reports z_a.

I characterize the Perfect Bayesian Equilibria of the game. a's decision is the choice to seek information from b, $c \in \{0, 1\}$. b's reporting strategy is $z_b : \{0, 1\} \times [0, 1] \to [0, 1] \cup \{\emptyset\}$. a's posterior belief about θ is given by $\mu : \{0, 1\} \times [0, 1] \times ([0, 1] \cup \{\emptyset\}) \to \Delta([0, 1])$ and her final report is $z_a : \{0, 1\} \times [0, 1] \times ([0, 1] \cup \{\emptyset\}) \to [0, 1]$.

A1.2 Notes on Model

This model is a straightforward application of Crawford and Sobel (1982) model of cheap talk in which the receiver (bureaucrat a) must pay a cost to access the sender's (bureaucrat b's) message. I choose the most common parameterizations, the state $\theta \sim U[0, 1]$ and quadratic utilities in order to concretely illustrate the mechanisms of interest. However, as shown by Crawford and Sobel (1982), the qualitative results hold under less restrictive parameterizations, namely when θ is distributed according to a continuous pdf on the interval $[0, 1]$ and utilities are twice continuously differentiable utility functions (Crawford and Sobel, 1982, p. 1443).

A1.3 Analysis

Consider first the report of the bureaucrat a. Her problem is:

$$\max_{z_a} E[-(\theta - z_a(z_b))^2 \mid z_b],$$

which yields the optimal report:

$$z_a^* = E[\theta \mid z_b].$$

So now we must find a's posterior belief about θ. Consider first the case in which the bureaucrat does not seek input ($c = 0$). In this case, the $E[\theta] = \frac{1}{2}$, which follows from the prior distribution of θ. In this case:

$$z_a^*(c = 0) = \frac{1}{2}.$$

Note further that $E[u_a(z_a, c = 0 \mid z_b, \theta)] = -\text{Var}[\theta] = -\frac{1}{12}$.

When $c = 1$, bureaucrat b can shape bureaucrat's a posterior belief through their report z_b. Crawford and Sobel (1982) provide characterization of the equilibrium reports, z_b.

Lemma 1 (Crawford and Sobel (1982))**.** *If $\beta > 0$, then there exists a partition of size N in which $N(\beta) = \left[-\frac{1}{2} + \frac{\sqrt{1+\frac{2}{\beta}}}{2}\right]$ where $[x]$ denotes the smallest integer greater than or equal to x, in which for all $j \in 0, ..., N$ partition boundaries are given by p_j, where:*

$$p_j = \frac{j}{N} + 2\beta j(j - N).$$

For θ in (p_j, p_{j+1}), b issues a report $z \in (p_j, p_{j+1})$. Upon observation of z_b, Bureaucrat a's posterior belief $\mu(z_b) \sim U[p_j, p_{j+1}]$. Bureaucrat a reports $E[\mu(z_b)] = \frac{p_j + p_{j+1}}{2}$.

If $\beta = 0$, then b issues report $z_b = \theta$. Bureaucrat a's posterior belief $\mu(z_b) = z_b$, and she reports $z_a = z_b$.

Proof: See Crawford and Sobel (1982).
Note further that in this part, bureaucrat a's expected utility is characterized in Lemma 2, which follows from the characterization of the variance under the partition equilibrium in Lemma 1, as characterized by Crawford and Sobel (1982).

Lemma 2. *If $\beta > 0$, bureaucrat a's expected utility from collaboration is given by:*

$$E[u_a(z_a, c = 1 \mid z_b, \theta)] = -\frac{1}{12N^2} - \frac{\beta^2(N^2 - 1)}{3} - c.$$

If $\beta = 0$, bureaucrat a's expected utility from collaboration is $-c$.

Proof: Given quadratic preferences, bureaucrat a's expected utility is:

$$-\text{Var}[\theta] - c.$$

When $\beta = 0$, bureaucrat a learns θ, so $\text{Var}[\theta] = 0$. When $\beta > 0$, Crawford and Sobel (1982) show that $\text{Var}[\theta] = \frac{1}{12N^2} + \frac{\beta^2(N^2-1)}{3}$. ■

Comparing bureaucrat a's expected utility from collaboration, Proposition 1 characterizes the Perfect Bayesian equilibria of the game.

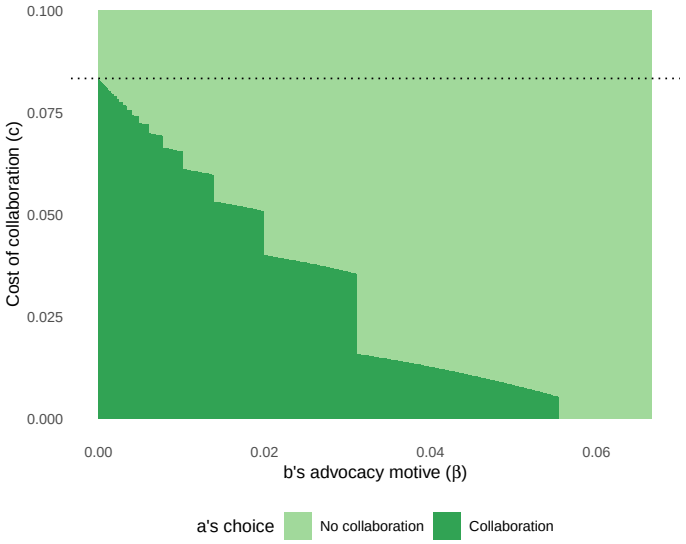

Figure A1: This figure plots regions that feature bureaucratic collaboration as a function of bureaucrat b's advocacy motives (β) and the cost of collaboration (c). The horizontal dotted line plots $c = \frac{1}{12}$, the maximum cost at which collaboration can be sustained when bureaucrats a and b prefer the same report (i.e., $\beta = 0$).

Proposition 1. *Bureaucrat a chooses to collaborate whenever:*

$$-\frac{1}{12N^2} - \frac{\beta^2(N^2-1)}{3} - c \geq -\frac{1}{12}.$$

The collaboration subgame proceeds according to Lemma 1. In the subgame without collaboration, bureaucrat a reports $z_a = \frac{1}{2}$, the mean of her prior over θ.

Proof: Follows from Lemmas 1-2 and the above discussion. ■

Figure A1 plots bureaucrat a's equilibrium collaboration decision. In the darker green region, the a opts to collaborate. The "jagged" boundary of the two regions reflects the partition structure of the equilibrium with collaboration. Above the dotted horizontal line ($c = \frac{1}{12}$), collaboration cannot be sustained even when both bureaucrats share the same preferences (i.e., $\beta = 0$). In the lighter region below that line, strategic reporting by b undermines the informational value (relative to the cost) of collaboration. Within this region, if b sincerely reported $\theta + \beta$, a could simply subtract β in order to learn and report θ. This illustrates the cost of strategic behavior.

A1.4 Empirical Implications

Empirically, this model informs predictions about two phenomena in the paper, as follows:

1. **Selection into collaboration**: Collaboration is:

   (a) Decreasing in the cost of collaboration c, which follows immediately from the inequality in Proposition (1).

(b) Piecewise decreasing in the difference in preferences between a and b, as parameterized by β. This is evident from two steps. First, note that N is decreasing in β. We can show this by evaluating the partial derivative of the continuous function from which N is realized (denote this function as N^*).

$$\frac{\partial N^*}{\partial \beta} = -\frac{1}{2\beta^2\sqrt{\frac{2+\beta}{\beta}}} < 0 \tag{1}$$

$$\frac{\partial \text{Var}[\theta \mid c=1]}{\partial N} = \tag{2}$$

$$\frac{\partial \text{Var}[\theta \mid c=1]}{\partial \beta} = \frac{2\beta(N^2-1)}{3} \geq 0 \tag{3}$$

2. **Effect of collaboration**: Collaboration results in more accurate data submitted by bureaucrat a. This follows by observing that collaboration implies that the inequality in Proposition 1 is satisfied. Because $c > 0$, this implies that:

$$\underbrace{\frac{1}{12N^2} - \frac{\beta^2(N^2-1)}{3}}_{\text{Var}[\theta|c=1]} < \underbrace{\frac{1}{12}}_{\text{Var}[\theta|c=0]} .$$

A2 Selection, balance, and attrition

The experimental sample consisted of 1,102 municipalities. It included all municipalities (excluding corregimientos) except the city of Bogotá. In sum, 902 municipalities completed the DT survey. The geographic distribution of municipalities is plotted in Figure A2. The map shows no evidence of geographic sorting of the municipalities that completed the DT survey.

Figure A3 summarizes differences in standardized baseline covariates between the experimental sample and municipalities that completed the DT survey (left). These differences are small in magnitude (all are $\leq \pm 0.05$ standard deviations) and do not reveal evidence of differential selection into completion of the DT survey. The right panel summarizes differences in the same covariates between the experimental sample and municipalities in the ultimate interview sample ($n = 30$). Consistent with the stratified random sampling strategy for interviews, interviewed municipalities are larger and more urban than the full sample of municipalities.

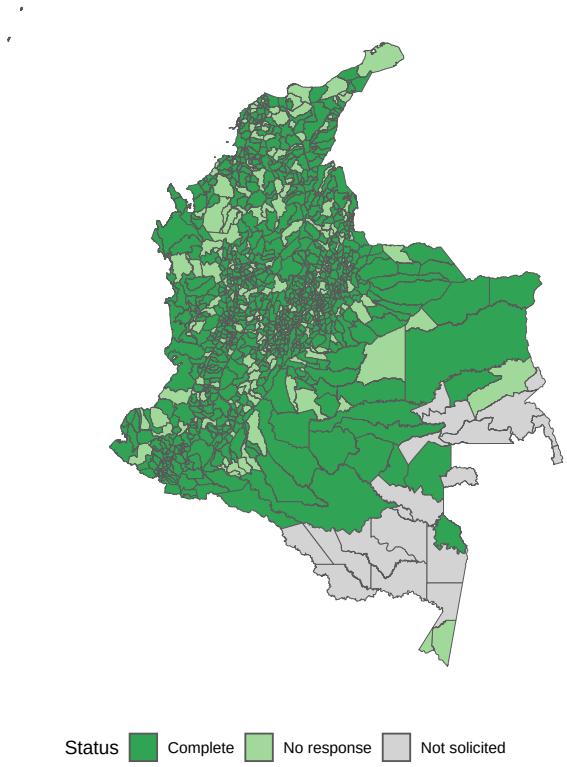

Figure A2: This map depicts the experimental sample of 1,102 municipalities and shades the 902 municipalities that completed the DT instrument in dark green.

Figure A4 assesses balance between the treatment and control groups in the encouragement experiment. The left panel provides little evidence of imbalance. Note that the experiment was blocked on municipal category so there should be (and is) minimal imbalance in treatment conditions for those units. In the right panel, we observe similar levels of balance in the municipalities that completed the DT survey.

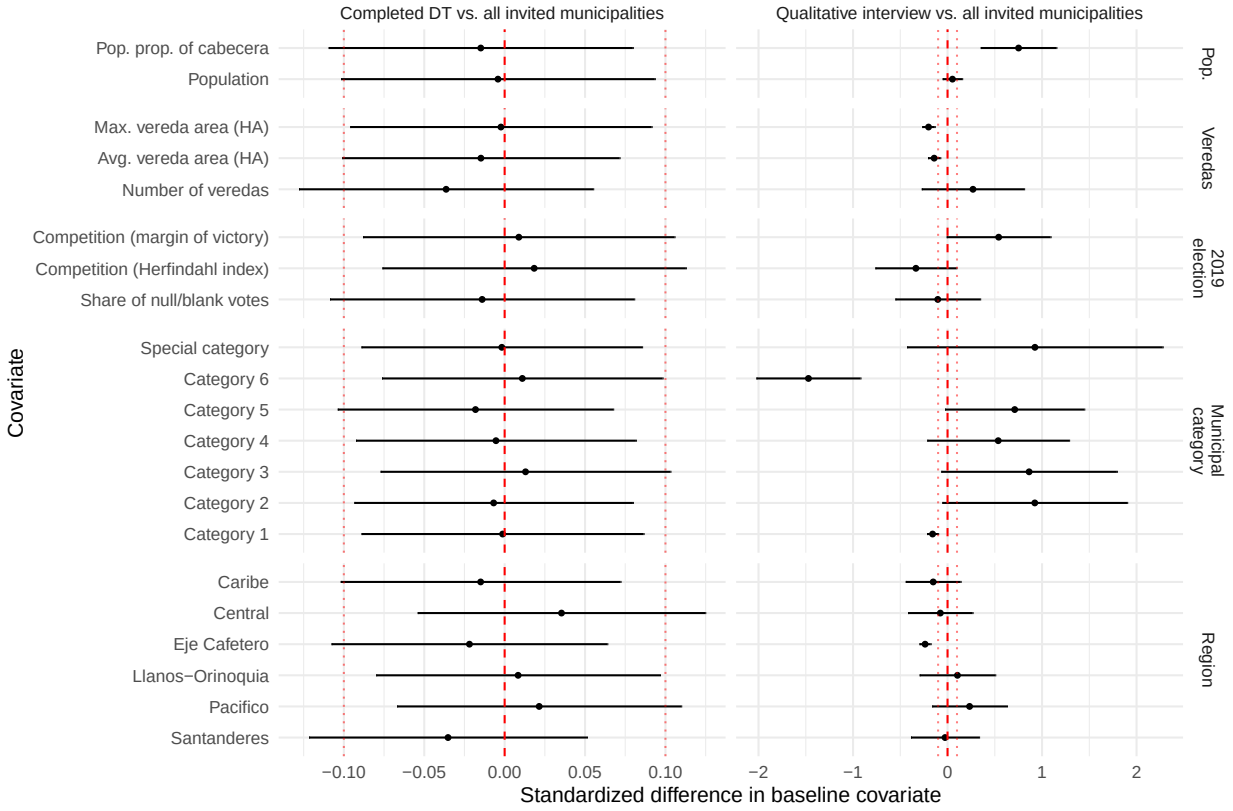

Figure A3: Standardized differences in a battery of municipal covariates between the experimental sample and the sample that completed the DT data collection (left) and between the experimental and qualitative interview samples (right).

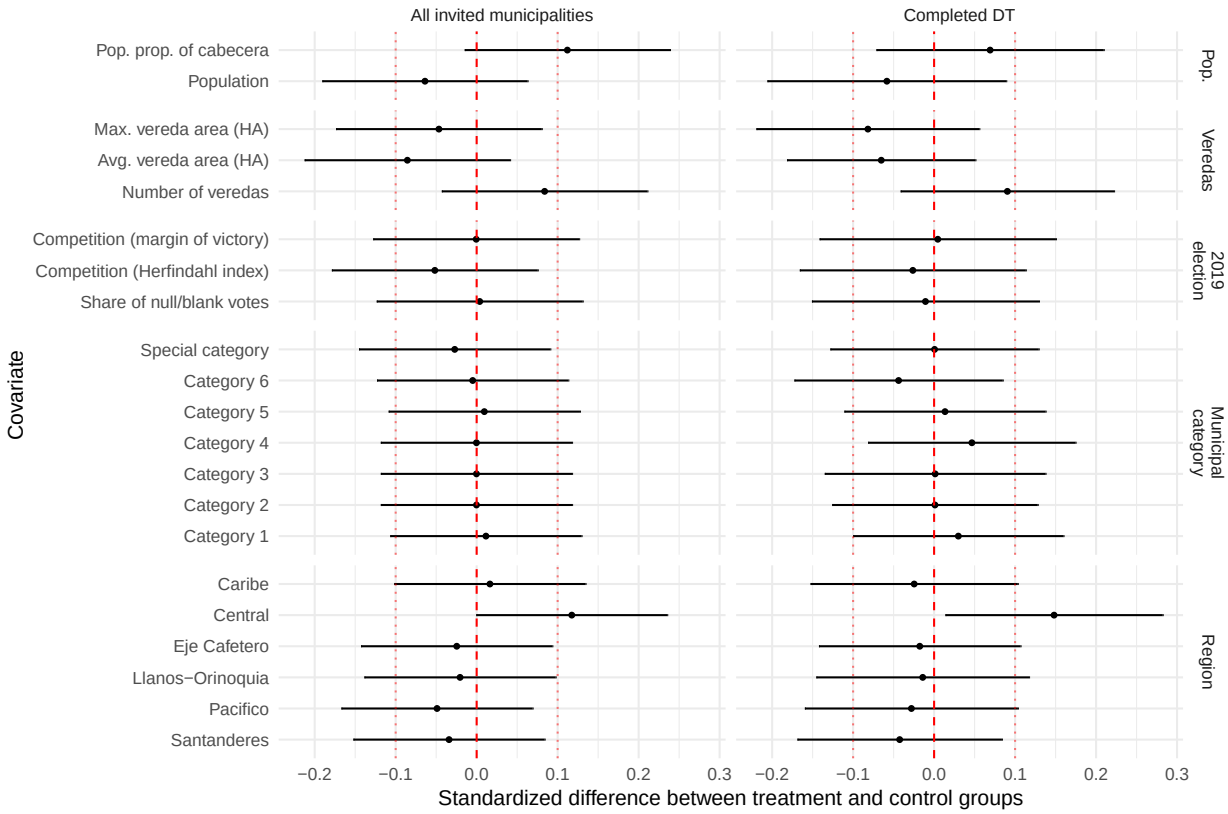

Figure A4: Standardized differences in a battery of municipal covariates between the treatment and control groups (left) and between treatment and control municipalities that completed the DT survey (right).

DT measures		External referent	
Measure	Operationalization	Source	Operationalization
Presence of marginalized groups	Binary indicators	2018 Census population counts by ethnic group	Binned to indicator when count ≥ 1
Count of a given natural disaster in last five years	Binned to an indicator when count ≥ 1	Indicator for a given natural disaster registered with the Unidad Nacional para la Gestión del Riesgo de Desastres between 2017 and 2021	Binary indicators
Count of urban and rural schools in municipality	Raw counts by zone	Counts of urban and rural schools according to Ministry of Education	Raw counts by zone.

Table A1: Description of DT survey items and external referents.

A3 Outcome measures

A3.1 Administrative data benchmarks and data validation

Many of the measures in the DT survey do not have readily available external referents and it is therefore difficult to assess quality.[1] Table A1 details the sources and (harmonized) operationalization of these indicators.

The external referents also provide an opportunity for benchmarking the accuracy of the DT data. In Figure A5, I use the census and natural disaster data to examine the agreement between binary indicators between DT and the external source. I compare this to the external source paired with a random draw from the distribution of the external source. The histograms plot the count of indicators in agreement. In both cases, the DT data agrees with the administrative referent at at a substantially higher rate than the random draws. The plots include Kolmogorov–Smirnov test statistics and p-values.

As a second means of validation, I take advantage of the fact that natural disasters are spatially concentrated and do not obey municipal boundaries. I look at the reported incidence of the same disaster in neighboring municipalities in both the DT data and the UNGRD administrative data. Figure A6 shows that in the DT survey data, neighbor municipalities of municipality i are more likely to have reported experiencing a disaster if i reports experiencing a disaster. These patterns are similar to—and often stronger than—analogous metrics from the UNGRD data. This suggests that the natural disaster reports capture the occurrence of commmon natural phenomena.

A3.2 Principal components index

I measure the content of reports using four principal components indices, one per section. This section reports information about the principal components used in the paper. Table A2 reports the number of variables (and thus principal components) for each section, as well as the proportion of variance explained by the first two principal components. Table A3 reports the correlation between the first principal components for each section. This is the orientation of the indices used in Figure 3. Table A4 reports the largest factor loadings (in absolute value) and the associated questions for each index. It reveals that the principal components indices are negatively correlated with log population and DNP's municipal performance index

[1] If more referents were available, the DT data collection would be largely redundant with other sources, raising other concerns.

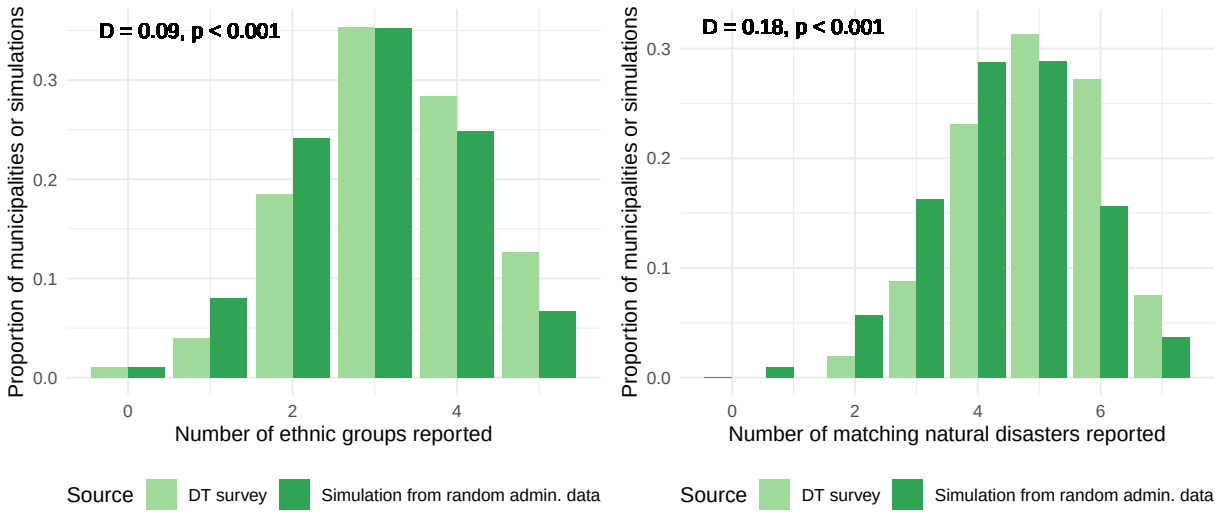

Figure A5: This figure reports a test of whether the agreement between the binary indicators from the DT data and external referent is greater than the agreement between the external referent and random draws from its distribution. For both sets of indicators, the DT data substantially outperforms the random sample. D denotes the Kolmogorov-Smirnov test statistic.

		Proportion of variance explained	
Index (DT survey section)	# of variables	First PC (analyzed)	Second PC (not analyzed)
Administrative capacity	40	0.083	0.063
Mobility	18	0.150	0.136
Infrastructure	56	0.080	0.043
Economic development	57	0.069	0.045

Table A2: Number of variables in each principal components index and the proportion of variance explained by the first two principal components.

(Table A5), two proxies of municipal development.

A4 Organigram data

One concern about the use of organigram data is that labels of units/secretariats are semantic. If this were the case, we would expect that some entities substitute for each other and measures of the organization of units, including the clusters, could be quite noisy or even misleading. To examine the empirical basis for this concern, I take indicators for the presence of each of 62 units and construct a correlation matrix. Table A6 reports the correlations greater than 0.3 in absolute value for any unit that appears in the organigram of at least 50 municipalities. Importantly, only two correlations are < -0.3, suggesting limited substitution of offices. There are many units that tend to emerge together, however. This suggests that some institutions may serve complementary functions.

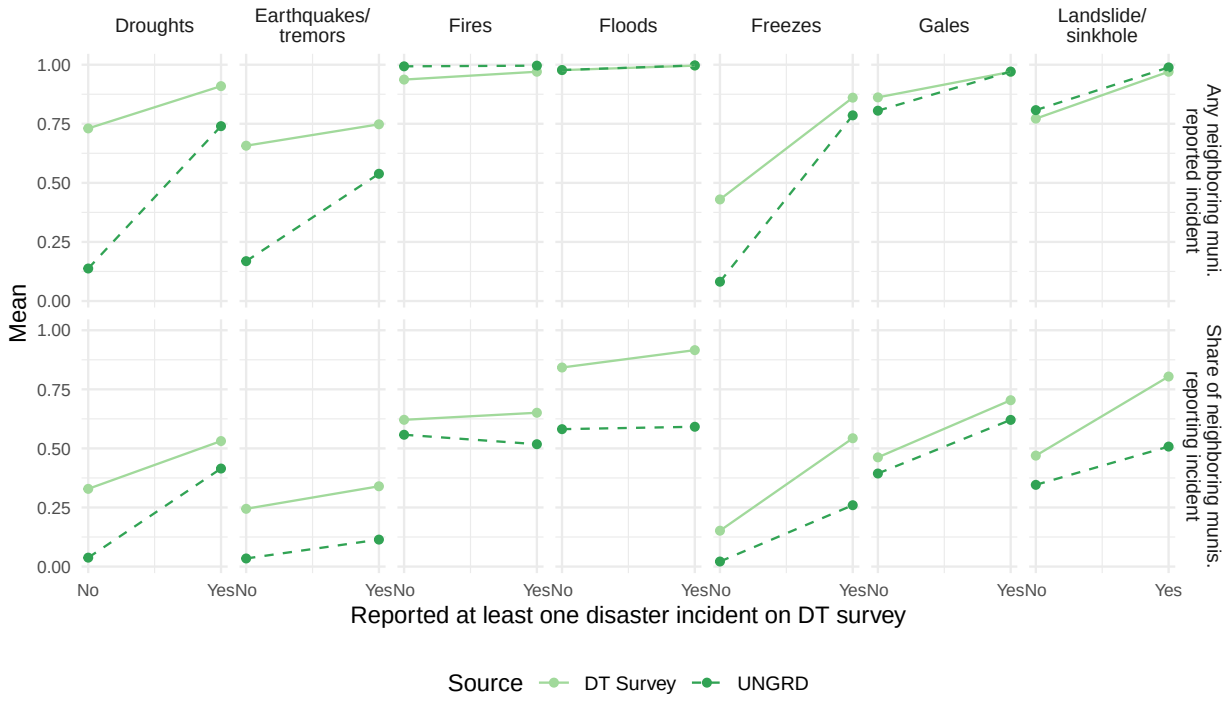

Figure A6: This figure examines two measures of the likelihood that a neighboring municipality reported experiencing a disaster as a function of whether a municipality experienced a disaster. The top row examines whether any neighbor experienced a disaster whereas the bottom row examines the share of neighboring municipalities that experienced a disaster.

	Admin. capacity	Mobility	Infrastructure	Econ. Dev.
Admin. capacity	(1)	0.036	0.533	0.300
Mobility	0.035	(1)	0.059	0.024
Infrastructure	0.533	0.059	(1)	0.480
Econ. Dev.	0.300	0.024	0.480	(1)

Table A3: Correlation matrix of first principal components utilized in the analysis.

Index	Factor loading	Item
AC	-0.284	Black/Afro-Colombians are present in municipality
AC	-0.263	Indigenous are present in municipality
AC	-0.251	Municipality has conducted employment training in last five years
AC	-0.240	Raizales are present in municipality
AC	-0.235	Palenqueros are present in municipality
M	0.436	Motorcycle is most common form of transportation to nearest village (vereda).
M	0.431	Motorcycle is the most common form of transportation to the furthest village (vereda)
M	-0.323	Leaving the center of the municipality, normally how much money does the majority of the population spend to get to the farthest village (vereda) in the municipality?
M	-0.305	Boat is the most common form of transportation to the nearest village (vereda)
M	-0.286	Boat is the most common form of transportation to the furthest village (vereda)
I	-0.281	There exist work training and human development institutions that are currently in use.
I	-0.266	There exist technical/vocational institutes that are currently in use.
I	-0.235	In the last five years, expansion of telephone lines has occurred in the municipality.
I	-0.233	There exist fire stations that are currently in use
I	-0.224	There exists infrastructure for transformation.
ED	-0.336	Schools in municipality source supplies locally (in municipality).
ED	-0.310	Programs for School Meals (*Programas de Alimentación* Escolar) source supplies locally (in municipality)
ED	-0.306	Health centers and hospitals source supplies locally (in municipality).
ED	-0.305	The army, police, or prisons and jails source supplies locally (in municipality).
ED	-0.297	Nutrition programs source supplies locally (in municipality).

Table A4: Five largest factor loadings (in absolute value) for the first principal component of each index.

Index	Log population	Municipal performance index (MDM)
Admin. capacity	-0.509	-0.107
Mobility	-0.092	-0.149
Infrastructure	-0.618	-0.240
Economic development	-0.371	-0.038
Aggregate index	-0.618	-0.202

Table A5: Correlation of principal components indices with municipal population (logged) and the DNP's municipal performance index.

Unit 1	Unit 2	Correlation	Proportion of municipalities with...			
			Neither	Only unit 1	Only unit 2	Both units
21- Secretaría General	12 - Secretaría de Gobierno	-0.61	0.10	0.50	0.00	0.09
15 - Secretaría de Infraestructura	10 - Secretaría de Planeación	-0.33	0.04	0.70	0.00	0.19
77 - Medio Ambiente / Gestión ambiental	63 - Vivienda	0.30	0.91	0.03	0.00	0.02
69 - Aseguramiento	52 - Archivo / Gestión documental	0.30	0.69	0.23	0.00	0.06
68 - Deportes y recreación	62 - Contratación	0.31	0.81	0.05	0.00	0.04
72 - Programas sociales	60 - Banco de proyectos	0.31	0.80	0.14	0.00	0.04
70 - Salud pública	52 - Archivo / Gestión documental	0.31	0.63	0.19	0.00	0.11
67 - Cultura y artes	54 - Talento humano	0.31	0.72	0.10	0.00	0.08
77 - Medio Ambiente / Gestión ambiental	67 - Cultura y artes	0.32	0.80	0.14	0.00	0.04
68 - Deportes y recreación	65 - Infraestructura y Obras Públicas	0.32	0.79	0.07	0.00	0.05
69 - Aseguramiento	60 - Banco de proyectos	0.32	0.79	0.13	0.00	0.05
70 - Salud pública	49 - Tesorería	0.33	0.66	0.16	0.00	0.10
67 - Cultura y artes	65 - Infraestructura y Obras Públicas	0.33	0.76	0.06	0.00	0.06
66 - Subsecretaría de educación	59 - Contabilidad	0.33	0.76	0.15	0.00	0.06
68 - Deportes y recreación	52 - Archivo / Gestión documental	0.34	0.66	0.20	0.00	0.09
15 - Secretaría de Infraestructura	13 - Secretaría de Educación	0.34	0.60	0.14	0.00	0.14
66 - Subsecretaría de educación	54 - Talento humano	0.35	0.78	0.13	0.00	0.06
57 - Presupuesto o finanzas públicas	49 - Tesorería	0.35	0.60	0.12	0.00	0.14
68 - Deportes y recreación	59 - Contabilidad	0.35	0.73	0.13	0.00	0.08
68 - Deportes y recreación	57 - Presupuesto o finanzas públicas	0.35	0.67	0.19	0.00	0.09
69 - Aseguramiento	54 - Talento humano	0.35	0.79	0.13	0.00	0.05
24 - Secretaría Jurídica	19 - Secretaría de Transporte	0.35	0.83	0.12	0.00	0.03
67 - Cultura y artes	52 - Archivo / Gestión documental	0.35	0.64	0.18	0.00	0.11
69 - Aseguramiento	57 - Presupuesto o finanzas públicas	0.35	0.70	0.22	0.00	0.06
65 - Infraestructura y Obras Públicas	54 - Talento humano	0.35	0.76	0.12	0.00	0.07
65 - Infraestructura y Obras Públicas	59 - Contabilidad	0.36	0.75	0.13	0.00	0.07
66 - Subsecretaría de educación	57 - Presupuesto o finanzas públicas	0.36	0.70	0.21	0.00	0.07
66 - Subsecretaría de educación	63 - Vivienda	0.36	0.89	0.02	0.00	0.03
54 - Talento humano	52 - Archivo / Gestión documental	0.36	0.64	0.18	0.00	0.12
70 - Salud pública	68 - Deportes y recreación	0.36	0.75	0.07	0.00	0.07
65 - Infraestructura y Obras Públicas	57 - Presupuesto o finanzas públicas	0.36	0.68	0.20	0.00	0.09
69 - Aseguramiento	59 - Contabilidad	0.36	0.77	0.15	0.00	0.06
67 - Cultura y artes	59 - Contabilidad	0.36	0.71	0.11	0.00	0.09
70 - Salud pública	60 - Banco de proyectos	0.37	0.73	0.09	0.00	0.09
60 - Banco de proyectos	54 - Talento humano	0.37	0.73	0.09	0.00	0.09

Unit 1	Unit 2	Correlation	Proportion of municipalities with...			
			Neither	Only unit 1	Only unit 2	Both units
53 - Inspección / Dirección de Policía	52 - Archivo / Gestión documental	0.37	0.37	0.04	0.00	0.26
70 - Salud pública	65 - Infraestructura y Obras Públicas	0.38	0.77	0.05	0.00	0.07
70 - Salud pública	67 - Cultura y artes	0.38	0.73	0.09	0.00	0.09
19 - Secretaría de Transporte	13 - Secretaría de Educación	0.39	0.68	0.17	0.00	0.10
67 - Cultura y artes	57 - Presupuesto o finanzas públicas	0.39	0.66	0.16	0.00	0.12
67 - Cultura y artes	60 - Banco de proyectos	0.39	0.73	0.09	0.00	0.09
68 - Deportes y recreación	60 - Banco de proyectos	0.39	0.76	0.10	0.00	0.08
68 - Deportes y recreación	66 - Subsecretaría de educación	0.40	0.82	0.04	0.00	0.05
59 - Contabilidad	49 - Tesorería	0.40	0.66	0.13	0.00	0.12
71 - Prestación de servicios de salud	70 - Salud pública	0.40	0.82	0.14	0.00	0.04
66 - Subsecretaría de educación	65 - Infraestructura y Obras Públicas	0.41	0.84	0.07	0.00	0.05
19 - Secretaría de Transporte	15 - Secretaría de Infraestructura	0.41	0.69	0.16	0.00	0.11
60 - Banco de proyectos	52 - Archivo / Gestión documental	0.42	0.65	0.17	0.00	0.12
71 - Prestación de servicios de salud	69 - Aseguramiento	0.42	0.91	0.05	0.00	0.02
59 - Contabilidad	52 - Archivo / Gestión documental	0.42	0.64	0.16	0.00	0.14
57 - Presupuesto o finanzas públicas	52 - Archivo / Gestión documental	0.43	0.59	0.12	0.00	0.17
70 - Salud pública	66 - Subsecretaría de educación	0.43	0.79	0.03	0.00	0.06
70 - Salud pública	59 - Contabilidad	0.43	0.72	0.10	0.00	0.10
78 - Adulto mayor	74 - Familias en acción	0.44	0.90	0.07	0.00	0.03
70 - Salud pública	57 - Presupuesto o finanzas públicas	0.44	0.66	0.16	0.00	0.13
67 - Cultura y artes	66 - Subsecretaría de educación	0.45	0.79	0.03	0.00	0.07
17 - Secretaría de Salud	13 - Secretaría de Educación	0.45	0.44	0.03	0.00	0.25
70 - Salud pública	54 - Talento humano	0.46	0.74	0.08	0.00	0.10
60 - Banco de proyectos	57 - Presupuesto o finanzas públicas	0.47	0.67	0.15	0.00	0.13
57 - Presupuesto o finanzas públicas	54 - Talento humano	0.48	0.67	0.05	0.00	0.14
60 - Banco de proyectos	59 - Contabilidad	0.50	0.73	0.09	0.00	0.11
70 - Salud pública	69 - Aseguramiento	0.54	0.81	0.01	0.00	0.07
59 - Contabilidad	54 - Talento humano	0.56	0.74	0.06	0.00	0.12
53 - Inspección / Dirección de Policía	50 - Comisaría de Familia	0.64	0.26	0.15	0.00	0.57
59 - Contabilidad	57 - Presupuesto o finanzas públicas	0.70	0.70	0.10	0.00	0.19
68 - Deportes y recreación	67 - Cultura y artes	0.76	0.81	0.05	0.00	0.13

Table A6: Correlation between organizational units within the same *alcaldía*. All institutions present in ≥ 50 municipalities with correlations $\geq |0.3|$ are reported.

A5 SIGEP (resumé) Data

SIGEP is the acronym for Sistema de Información y Gestión del Empleo Público (roughly Public Employment Information and Management System). I utilize two components in the SIGEP data. First, public sector workers submit resumés (*hojas de vida*) with relevant public sector experience. Public sector workers submit their own information to this portion of the system. Second, individual bureaucratic entities (e.g., an *alcaldía*) enter their salary scales, by type and title.

Search parameters: I have the reported first and last names of each agent that reported data for DT. I conducted up to two searches in the SIGEP database for each individual:[2]

- Search #1: [apellido #1] [apellido #2]
- Search #2 (if individual was not located in Search #1): [nombre #1] [apellido #1]

If there were multiple possible candidates with a name, I used the entity name to select the correct individual and CV.

Selecting Municipalities for Wage Data: Since public sector wages are constrained by municipal category, I gather wage data for a random sample of *alcaldías*, stratifying on municipal category.

A5.1 Data Validation

I examine the overlap between reported tenures of reporters in the DT survey and SIGEP data. The search procedure above located 277 of 902 officials, of which 224 provide complete information on their position. Among these 224 complete observations, 71% fall into the same bin in both data sources and 92% fall within a margin of one bin (of five) on either side.
I plot the distribution of tenures in the *alcaldía*, for directors (appointees) and non-directors (career civil servants) in Figure A7. This graph shows that across data sources, directors have substantially shorter tenures on average, with the majority appointed within the current mayor's term (first two categories). The non-directors in the SIGEP data have longer tenures, on average, than the self reports in the DT data. This is likely driven by the fact that longer-term employees have had more opportunities to enter their data in SIGEP.

A5.2 Hierarchy within bureaucratic subunits

Figure A8 plots wage ranges for each position/category of public employee. There are five broad categories: leadership (*director*), professional, technical, support (*asistencial*), and advisors (*asesor*). Since advisors are not employed in all municipalities, I focus on the first four categories that are common across almost all municipalities in the sample. To account for varying salaries, I plot wage ranges as a ratio of pay maximum pay for a position in the leadership rank (*director*). Formally, suppose that in municipality m in category c, the maximum wage for an employee in position p is denoted $\overline{w}_{pmc}$ and the minimum wage for such an employee is denoted $\underline{w}_{pmc}$. Denote a position in leadership as $p = D$. For each municipality, I thus construct measures:

$$\underline{R}_{pmc} = \frac{\underline{w}_{pmc}}{\overline{w}_{Dmc}} \qquad\qquad \overline{R}_{pmc} = \frac{\overline{w}_{pmc}}{\overline{w}_{Dmc}}$$

[2]Most bureaucrats reported two first names ("nombres") and two last names ("apellidos") consistent with Colombian naming customs.

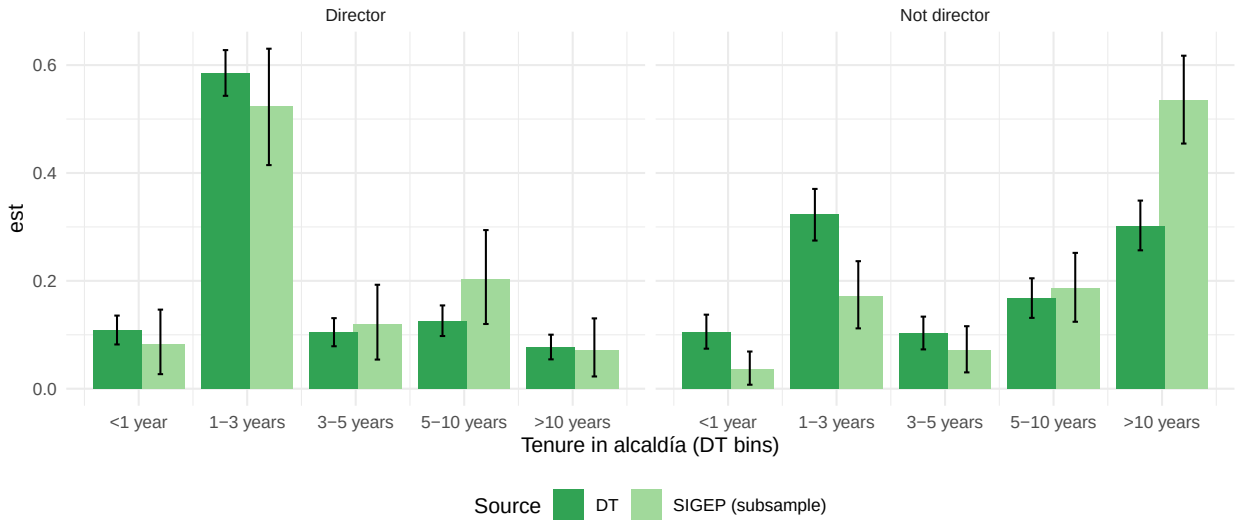

Figure A7: The distribution of tenure between directors (appointees) and non-directors (career civil servants) in the DT data and the subsample of employees in SIGEP.

Clearly $\overline{R}_{Dmc} = 1$ by construction. Figure A8 plots the average ranges within each category, c. This figure supports the claim of the rank and value of leadership positions in the *alcaldías*. While there does exist some evidence of wage compression between the highest-paid professionals and leadership in category 6 municipalities, bureaucrats in leadership earn more than individuals in all other positions.
Figure A9 shows the duration of a bureaucrat's current appointment and tenure in the *alcaldía* (across roles) as a function of whether they are an executive versus non-executive. It reveals that executives have substantially shorter tenures than civil servants. Their average tenure in their role is approximately one-third of the average tenure of civil servants. Moreover, most executives were appointed by the current mayor whereas most civil servants were not.

A6 Intent-to-treat Analysis

Table A7 reports the reduced-form intent-to-treat effects from the encouragement experiment. While estimates are notably small and imprecisely estimated, five of the signs match the signs from the OLS analysis. Recall from Table 4 that the first-stage effect of encouragement (with the binary collaboration measure) is 0.043. One interpretation of the ratio of $\widehat{ITT}$ to $\widehat{\beta_{OLS}}$ is that if: (1) the association between collaboration and data quality were entirely the causal effect of collaboration; and (2) causal effects were homogenous across *alcaldías*, the ratio should match the first-stage.

A7 Organizational features and selection

Figure A10 reports the adjusted R^2 from a series of bivariate regressions of collaboration (three measures) on a series of municipal and bureaucrat-level characteristics. It shows that the secretary indicator (and the alternate leader coding) are the strongest predictors of collaboration. However even these measures have adjusted R^2's between 0.023 and 0.05, depending on the outcome.

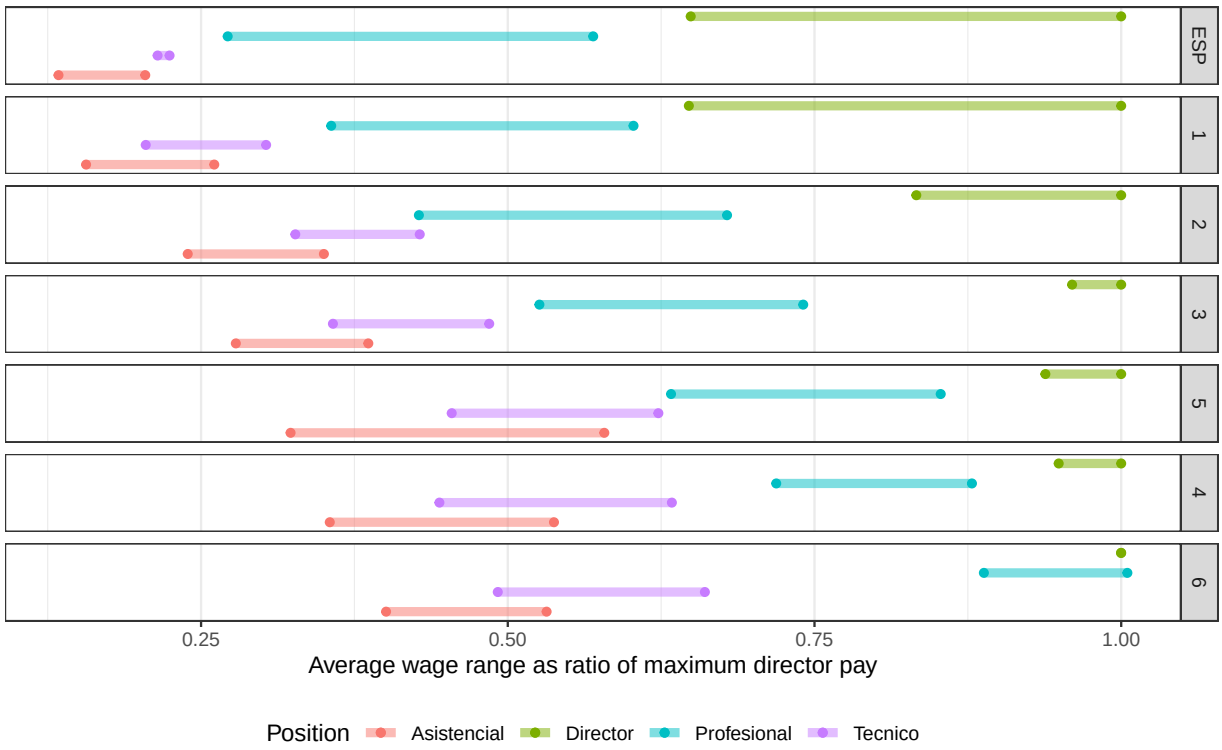

Figure A8: This figure plots the average range of wages for each employment category as a ratio of the maximum pay for directors by municipal category.

	Certainty (1)	Missingness† (2)	Accuracy (3)	Principal Components (4)	Duration† (5)	Enumerator evaluation (6)
PANEL I: INTENT-TO-TREAT (REDUCED FORM) ESTIMATES						
Encouragement	0.073	-0.013	0.010	-0.049	0.023	0.058
	(0.065)	(0.065)	(0.068)	(0.059)	(0.065)	(0.063)
PANEL II: OLS ESTIMATES						
Collaboration (Extensive margin)	0.532***	0.314*	0.184	-0.037	0.240*	0.457***
	(0.115)	(0.125)	(0.136)	(0.101)	(0.116)	(0.104)
$\widehat{ITT}/\widehat{\beta_{OLS}}$	0.137	-0.0132	0.053	1.301	0.095	0.126
Benchmark covariates	✓	✓	✓	✓	✓	✓
Enumerator FE					✓	✓
Outcome mean	0	0	0.037	1	-0.002	0
Outcome standard deviation	1	1	0.935	1	1.013	1
Number of Observations	902	902	846	902	875	902

$^{+}p < 0.1;^{*}p < 0.05;^{**}p < 0.01$

Table A7: Intent-to-treat effect estimates (Panel A) and reduced-form OLS estimates (Panel B) for the seven outcome indices in Figure 3. † indicates that the outcome measure is reverse-coded (so positive estimates indicate less missingness or shorter durations). Heteroskedasticity-robust standard errors in parentheses.

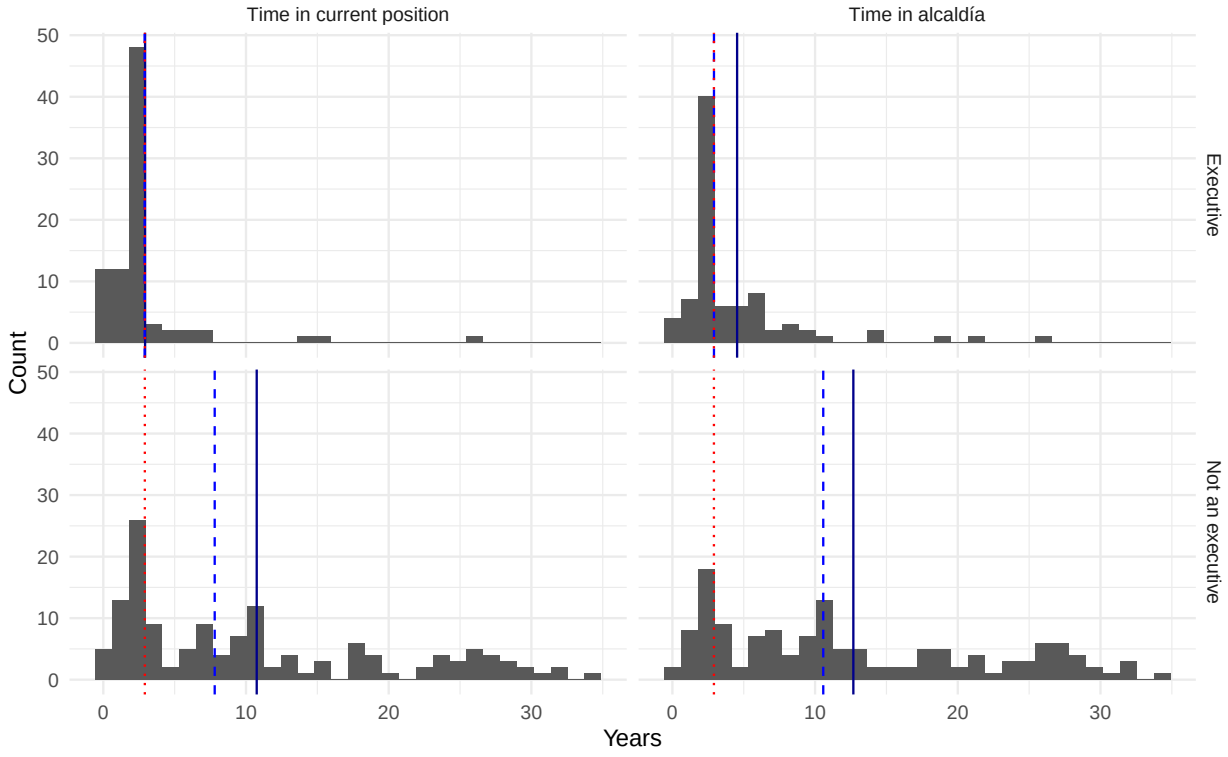

Figure A9: This figure plots the duration of a bureaucrat's (a) time in their current role; and (b) total time in the *alcaldía* as a civil servant or contractor (columns), as a function of whether they are an executive versus non-executive (rows). The dark blue lines give the mean (solid) and median (dashed) of the distributions. The red dotted denotes the tenure of the mayor (at the time of the data collection).

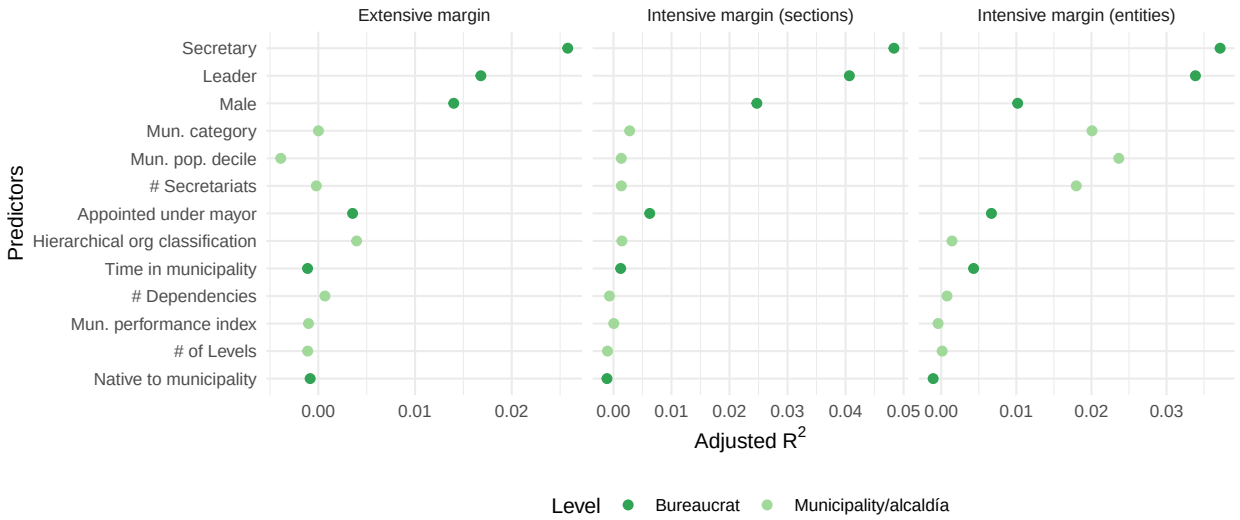

Figure A10: This figure plots the adjusted R^2 from bivariate regressions of collaboration on a series of municipal and bureaucrat-level characteristics.

Figure A11 reveals that the comparison of leaders to non-leaders results in virtually identical differences in selection into collaboration. The leader vs. non-leader categorization is broader but slightly more precise, as the precise titles within non-secretariat dependencies are somewhat more varied.

Figure A12 disaggregates the results in Figure 4 by whether a respondent was an invitee by PS or a replacement. The difference in rates of collaboration between invitee secretaries vs. non-secretaries is similar to the difference between replacement secretaries vs. non-secretaries. This suggests that the observed difference is not solely driven by the (somewhat ambiguous) selection process PS used to select municipal bureaucrats as reporters.

A8 Deviations from Pre-Analysis Plan

The pre-analysis plan sought to analyze the experiment as an encouragement design and proposed to estimate the LATE among compliers (those induced to collaborate by the treatment) and related IV estimands (for the continuous measures of collaboration). Obviously there was not a sufficiently strong first stage to pursue this design. In this sense, the paper goes far beyond the pre-analysis plan to understand the observational relationship between collaboration and the reported data. However, there are a few smaller departures from the PAP that are worth articulating.

1. As noted on page 16, I normalize both intensive margin measures of collaboration such that all measures of collaboration range from 0 to 1. For the sections measure, the normalization trivially rescales the coefficients. For the entities measure, I divide by the number of entities that respondents reported were present in the *alcaldía*. This addresses the issue that more populous municipalities have more complex organizations so the pre-specified count would have picked up both population and collaboration.

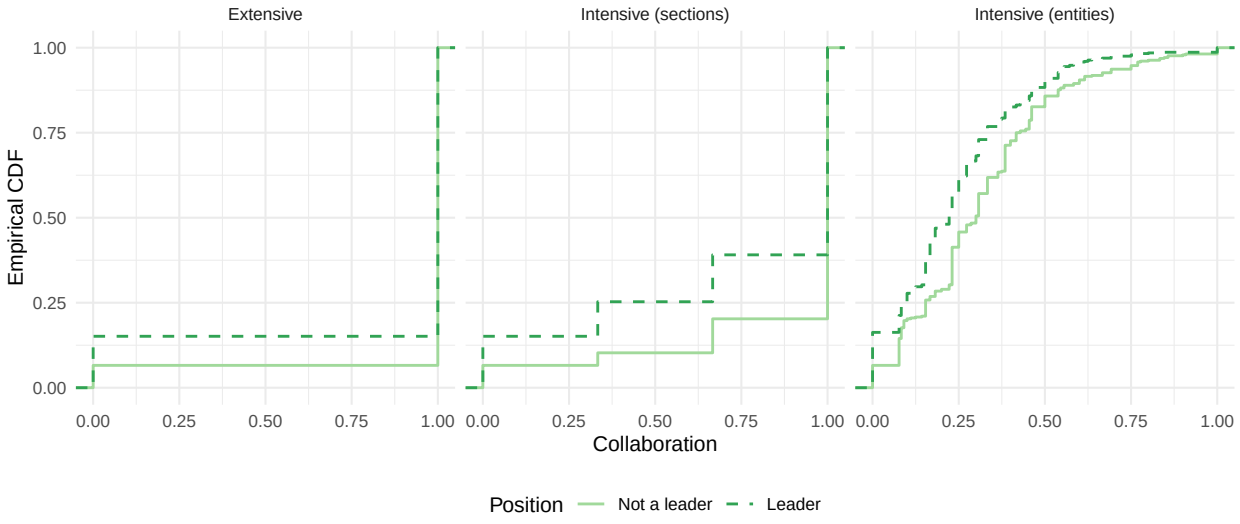

Figure A11: This figure replicates 4 but compares leaders to non-leaders rather than secretaries to non-secretaries.

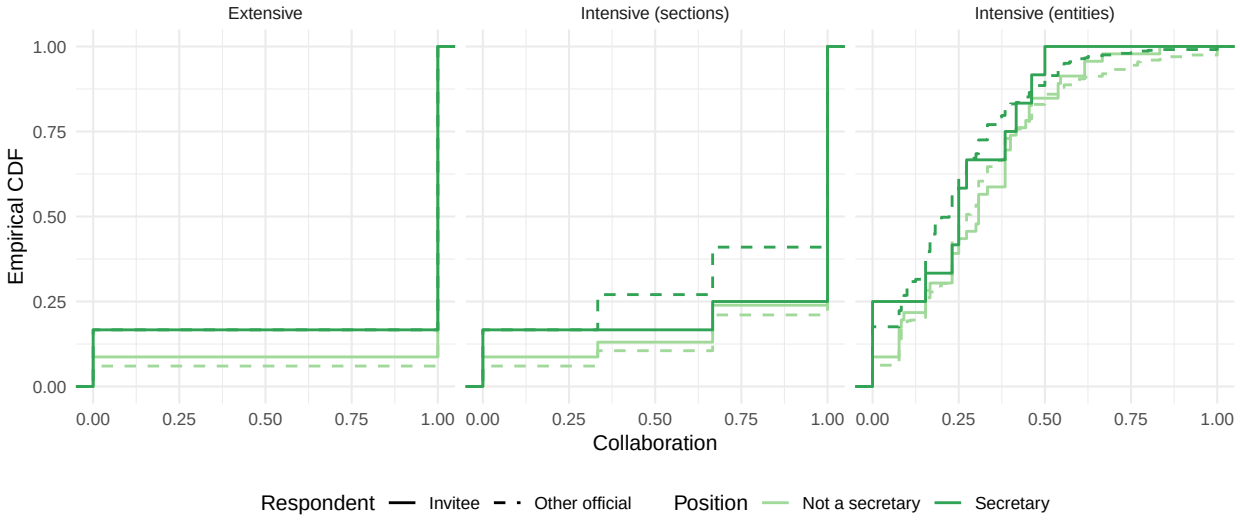

Figure A12: Disaggregating the ECDF in Figure 4 by whether the respondent was an initial invitee or a replacement. In both subsamples, secretaries are less likely to collaborate than non-secretaries.

2. In the principal components analysis, I focus on the first principal component of each section. I had prespecified two possible treatments of the principal components: (1) selection of all principal components with eigenvalues over 1; and (2) selection of the principal components that cumulatively explained 95% of the variation in the outcome. Neither approach reduced dimensionality sufficiently. For example, for the institutional capacity section, there were 40 variables (without conditionality in response). Fourteen (14) principal components had an eigenvalue over 1 and it required 36 principal components to explain the 95% of the variation in the data. Given the limited reduction in dimensionality under either approach, I focus on the first principal component for each section of the survey.

3. I analyze the duration of responses as an additional meta-data outcome, as noted in the text. At the time that the PAP was filed, I was not aware that this meta-data would be recorded.

Supplementary Appendix: References

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50 (6): 1431–1451.