The Limits of Decentralized Data Collection: Experimental Evidence from Colombia

Tara Slough—NYU and Natalia Garbiras-Díaz—EUI

September 2022

State information gathering

- "Statistics" famously derives from the word "state."
- Modern states collect vast amounts of data for use in policymaking and public administration.
  - The classics: censuses, vital statistics, land cadasters
  - Increasingly, methods of passive surveillance/monitoring
  - Data produced, submitted by bureaucrats → **this talk!**

State information gathering

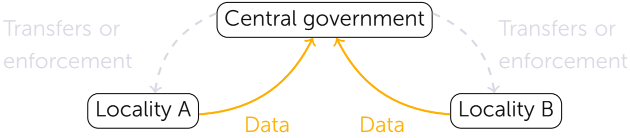
- "Statistics" famously derives from the word "state."
- Modern states collect vast amounts of data for use in policymaking and public administration.
  - The classics: censuses, vital statistics, land cadasters
  - Increasingly, methods of passive surveillance/monitoring
  - Data produced, submitted by bureaucrats → **this talk!**
- Policymakers and donors advance "data-driven governance" as a means to reduce corruption, waste, and inefficiency.
  - Goal: Data inputs should be used to affect policy.

State information gathering

- "Statistics" famously derives from the word "state."
- Modern states collect vast amounts of data for use in policymaking and public administration.
    - The classics: censuses, vital statistics, land cadasters
    - Increasingly, methods of passive surveillance/monitoring
    - Data produced, submitted by bureaucrats → **this talk!**
- Policymakers and donors advance "data-driven governance" as a means to reduce corruption, waste, and inefficiency.
    - Goal: Data inputs should be used to affect policy.

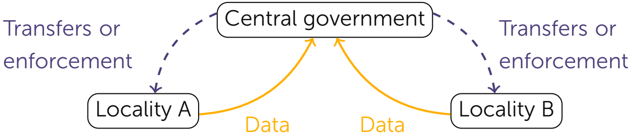
Decentralized data production

- Our focus: decentralized data production
  - Origin of many administrative datasets
  - Data requests from central to decentralized bureaucracies.



Decentralized data production

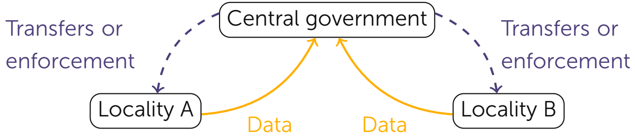
- Our focus: decentralized data production
  - Origin of many administrative datasets
  - Data requests from **central** to **decentralized** bureaucracies.



- Is this really a thing?
  - Local bureaucrats complain about these requests.
  - **Time-use surveys** of local bureaucrats suggest that these requests are time consuming Kalaj, Rogger, and Somani (2022)
  - National bureaucrats **seek advice** on eliciting complete, honest reports.

Decentralized data production

- Our focus: decentralized data production
  - Origin of many administrative datasets
  - Data requests from **central** to **decentralized** bureaucracies.



- Is this really a thing?
  - Local bureaucrats complain about these requests.
  - **Time-use surveys** of local bureaucrats suggest that these requests are time consuming Kalaj, Rogger, and Somani (2022)
  - National bureaucrats **seek advice** on eliciting complete, honest reports.

Overview

Framework: Simple decision theoretic model maps the behavior of decentralized bureaucrats onto statistical measurement framework.

- Bureaucrats choose how much **effort** to exert and whether to purposefully **distort** their reports to the central government.

Overview

Framework: Decision theoretic model that maps the behavior of decentralized bureaucrats onto statistical measurement framework.

Design: A field experiment in collaboration with the Colombian Procuraduría General de la Nación on the annual national transparency index (ITA).

- Manipulates of the visibility of PGN in using the data → measure the responses of bureaucrats in decentralized entities to increased oversight.
- An independent audit of a subset of items allows us to describe data quality.

Overview

Framework: Decision theoretic model that maps the behavior of decentralized bureaucrats onto statistical measurement framework.

Design: A field experiment in collaboration with the Colombian Attorney Inspector General's office on the annual national transparency index (ITA).

Findings:

- Reporting behavior responds to the visibility of oversight in both **selection into reporting** and the **information reported**.
- The audit suggests that three pathologies of measurement: **selection**, **intentional distortions**, and the **random error** in reports covary with true quality.

Related Literature

- State information about individual citizens like censuses and vital statistics

  (Scott, 1998; Lee and Zhang, 2017; Bowles, 2020; Sánchez-Talanquer, 2020)

- Administrative data quality usually in two settings:
  - Autocratic regimes, esp. with respect to economic data (Guriev and Treisman, 2019; Martínez, 2021; Trinh, 2021; Lorentzen, 2014; Wallace, 2016; Edmond, 2013)
  - Police data quality in the US (Eckhouse, 2022; Cook and Fortunato, 2022)

Related Literature

- State information about individual citizens like censuses and vital statistics (Scott, 1998; Lee and Zhang, 2017; Bowles, 2020; Sánchez-Talanquer, 2020)
- Administrative data quality usually in two settings:
  - Autocratic regimes, esp. with respect to economic data (Guriev and Treisman, 2019; Martínez, 2021; Trinh, 2021; Lorentzen, 2014; Wallace, 2016; Edmond, 2013)
  - Police data quality in the US (Eckhouse, 2022; Cook and Fortunato, 2022)
- Application: collection of data on corruption and transparency of state institutions (Ferraz and Finan, 2008; Larreguy, Marshall, and Snyder, 2018)

Related Literature

- State information about individual citizens like censuses and vital statistics (Scott, 1998; Lee and Zhang, 2017; Bowles, 2020; Sánchez-Talanquer, 2020)
- Administrative data quality usually in two settings:
  - Autocratic regimes, esp. with respect to economic data (Guriev and Treisman, 2019; Martínez, 2021; Trinh, 2021; Lorentzen, 2014; Wallace, 2016; Edmond, 2013)
  - Police data quality in the US (Eckhouse, 2022; Cook and Fortunato, 2022)
- Application: collection of data on corruption and transparency of state institutions (Ferraz and Finan, 2008; Larreguy, Marshall, and Snyder, 2018)

Theoretical Framework

The output: administrative data

- Central government asks decentralized entities to report some quantity, θ, to the central government.
  - E.g., public service outputs, budget execution, or transparency practices.
- A bureaucrat (or office) in the entity chooses whether to report, r.
  - If they do not make a report, $r = \emptyset$.
  - If they do report, $r \in \mathbb{R}$.

The output: administrative data

- Central government asks decentralized entities to report some quantity, θ, to the central government.
    - E.g., public service outputs, budget execution, or transparency practices.
- A bureaucrat (or office) in the entity chooses whether to report, r.
    - If they do not make a report, $r = \varnothing$.
    - If they do report, $r \in \mathbb{R}$.
- Conditional on reporting, r and θ may be different due to:
    - Intentional distortion, d.
    - Unintentional errors/random error, $\varepsilon \sim f(\cdot)$, where $f(\cdot)$ is a mean-zero density.

The output: administrative data

- Central government asks decentralized entities to report some **quantity, θ,** to the central government.
    - E.g., public service outputs, budget execution, or transparency practices.
- A bureaucrat (or office) in the entity chooses **whether to report, r**.
    - If they do not make a report, $r = \varnothing$.
    - If they do report, $r \in \mathbb{R}$.
- Conditional on reporting, r and θ may be different due to:
    - Intentional distortion, d.
    - Unintentional errors/random error, $\varepsilon \sim f(\cdot)$, where $f(\cdot)$ is a mean-zero density.

Data outputs

- Maps onto standard formulations of measurement error in statistics (e.g., Cochran, 1968; Rubin, 1976)
- The central government wants to know θ but observes r, which can suffer from:
  - Missingness when the bureaucrat does not report $r = \varnothing$.
  - Measurement error due to:
    - Intentional distortions (d)
    - Unintentional distortions (ε)

$$\underbrace{r}_{\text{Report}} = \begin{cases} \underbrace{\theta}_{\text{Truth}} + \underbrace{d}_{\substack{\text{Intentional} \\ \text{misreporting}}} + \underbrace{\varepsilon}_{\substack{\text{Random} \\ \text{error}}} & \text{if bureaucrat reports} \\ \varnothing & \text{otherwise} \end{cases}$$

Data outputs

- Maps onto standard formulations of measurement error in statistics (e.g., Cochran, 1968; Rubin, 1976)
- The central government wants to know θ but observes r, which can suffer from:
  - **Missingness** when the bureaucrat does not report $r = \varnothing$.
  - **Measurement error** due to:
    - Intentional distortions (d)
    - Unintentional distortions (ε)

$$\underbrace{r}_{\text{Report}} = \begin{cases} \underbrace{\theta}_{\text{Truth}} + \underbrace{d}_{\substack{\text{Intentional}\\ \text{misreporting}}} + \underbrace{\varepsilon}_{\substack{\text{Random}\\ \text{error}}} & \text{if bureaucrat reports} \\ \varnothing & \text{otherwise} \end{cases}$$

Central government data use

- Central government may use data to:
  - Target resources: **carrots**.
  - Target oversight or enforcement: **sticks** → **our empirical setting**.
- Key assumption: bureaucrats internalize (to some degree) entity outcomes.

Central government data use

- Central government may use data to:
  - Target resources: **carrots**.
  - Target oversight or enforcement: **sticks** → **our empirical setting**.
- Key assumption: bureaucrats internalize (to some degree) entity outcomes.
- Two policy instruments by the central government in enforcement setting → exogenous in the experiment.
  - Reliance on observed data to target oversight with probability $\rho(r) \in (0,1)$.
  - Penalties imposed when oversight reveals poor outcomes (θ) or reporting behavior (r): $P(\theta, r) > 0$.

Central government data use

- Central government may use data to:
  - Target resources: **carrots**.
  - Target oversight or enforcement: **sticks** → **our empirical setting**.
- Key assumption: bureaucrats internalize (to some degree) entity outcomes.
- Two policy instruments by the central government in enforcement setting → exogenous in the experiment.
  - Reliance on observed data to target oversight with probability $\rho(r) \in (0,1)$.
  - Penalties imposed when oversight reveals poor outcomes (θ) or reporting behavior (r): $P(\theta, r) > 0$.
- Governments struggle to set these policies, so much so that they are (sometimes) willing to randomize!

Central government data use

- Central government may use data to:
  - Target resources: **carrots**.
  - Target oversight or enforcement: **sticks** $\rightarrow$ **our empirical setting**.
- Key assumption: bureaucrats internalize (to some degree) entity outcomes.
- Two policy instruments by the central government in enforcement setting $\rightarrow$ exogenous in the experiment.
  - Reliance on observed data to target oversight with probability $\rho(r) \in (0,1)$.
  - Penalties imposed when oversight reveals poor outcomes (θ) or reporting behavior (r): $P(\theta, r) > 0$.
- Governments struggle to set these policies, so much so that they are (sometimes) willing to randomize!

What are bureaucrats doing?

- When reporting, bureaucrats exert effort and choose intentional distortions
  - Effort, $e \geq 0 \rightarrow$
    - Missingness: when $e = 0$.
    - Unintentional distortions: $\uparrow e \rightarrow \downarrow \text{Var}(\epsilon)$.
  - Intentional distortions, d
- Bureaucrat's utility:

$$U_B = \underbrace{\underbrace{-\rho(r)}_{\text{Pr(Audit)}} \; \underbrace{P(\theta, r)}_{\text{Penalty}}}_{\text{Oversight}} - \underbrace{c(e)}_{\text{Cost of effort}}$$

- Note: bureaucrats may not know precisely how oversight is exercised.

What are bureaucrats doing?

- When reporting, bureaucrats exert effort and choose intentional distortions
  - Effort, $e \geq 0 \rightarrow$
    - Missingness: when $e = 0$.
    - Unintentional distortions: $\uparrow e \rightarrow \downarrow \text{Var}(\epsilon)$.
  - Intentional distortions, d
- Bureaucrat's utility:

$$U_B = \underbrace{\underbrace{-\rho(r)}_{\text{Pr(Audit)}} \; \underbrace{P(\theta, r)}_{\text{Penalty}}}_{\text{Oversight}} - \underbrace{c(e)}_{\text{Cost of effort}}$$

- Note: bureaucrats may not know precisely how oversight is exercised.

Learning about the observed administrative data

- An exogenous increase in (perceived) oversight:
  - Should ↑ **rates of reporting**.
  - Changes the aggregate distribution of reports, conditional on reporting, by:
    1. ↑ Entities reporting;
    2. Changing incentives for intentional misreporting;
    3. ↓ Unintentional errors.
- Observation of the joint distribution of quality, θ, and reported quality, r, allow for learning about:
  - Bureaucrats' reporting behavior, including both effort and intentional distortions.
  - Bureaucrats' expectations about oversight by the central government.

Learning about the observed administrative data

- An exogenous increase in (perceived) oversight:
  - Should ↑ **rates of reporting**.
  - Changes the aggregate distribution of reports, conditional on reporting, by:
    1. ↑ Entities reporting;
    2. Changing incentives for intentional misreporting;
    3. ↓ Unintentional errors.
- Observation of the joint distribution of quality, θ, and reported quality, r, allow for learning about:
  - Bureaucrats' reporting behavior, including both effort and intentional distortions.
  - Bureaucrats' expectations about oversight by the central government.

Case Context

Our partner: The PGN

- The Procuraduría General de la Nación (PGN), is the principal watchdog agency in Colombia.
- PGN is implementing of the National Transparency Law of 2014.
  - This law mandated the creation of the National Transparency Index (ITA), the measure that we study.

Our partner: The PGN

- The Procuraduría General de la Nación (PGN), is the principal watchdog agency in Colombia.
- PGN is implementing of the National Transparency Law of 2014.
  - This law mandated the creation of the National Transparency Index (ITA), the measure that we study.
- PGN is also the principal user of the ITA data, as part of its preventative mandate.
  - This mandate seeks to prevent corruption or other public misconduct by monitoring of public officials and entities.

Our partner: The PGN

- The Procuraduría General de la Nación (PGN), is the principal watchdog agency in Colombia.
- PGN is implementing of the National Transparency Law of 2014.
  - This law mandated the creation of the National Transparency Index (ITA), the measure that we study.
- PGN is also the principal user of the ITA data, as part of its preventative mandate.
  - This mandate seeks to prevent corruption or other public misconduct by monitoring of public officials and entities.

The National Transparency Index (ITA)

- The Indice de Transparencia y Acceso a la Información, ITA, was first implemented in 2018.
  - We study the production of the 2020 index.
- $\approx 50k$ entities to report their compliance with ≈ 200 transparency practices.
  - All items are binary (yes/no).
  - Weighted to generate the 100-point ITA.
  - Used in PGN's preventative actions to guide monitoring/investigation.

The National Transparency Index (ITA)

- The Indice de Transparencia y Acceso a la Información, ITA, was first implemented in 2018.
  - We study the production of the 2020 index.
- $\approx 50k$ entities to report their compliance with ≈ 200 transparency practices.
  - All items are binary (yes/no).
  - Weighted to generate the 100-point ITA.
  - Used in PGN's preventative actions to guide monitoring/investigation.
- Unit of measurement, the entity, classified as:
  1. Traditional public sector entities including public entities, oversight bodies, and state-owned companies $\rightarrow$ **our focus**.
  2. Non-traditional entities are persons or legal entities that contract with the state to provide public services or manage public funds.
  3. Political parties or social movements.

The National Transparency Index (ITA)

- The Indice de Transparencia y Acceso a la Información, ITA, was first implemented in 2018.
  - We study the production of the 2020 index.
- $\approx 50k$ entities to report their compliance with ≈ 200 transparency practices.
  - All items are binary (yes/no).
  - Weighted to generate the 100-point ITA.
  - Used in PGN's preventative actions to guide monitoring/investigation.
- Unit of measurement, the entity, classified as:
  1. Traditional public sector entities including public entities, oversight bodies, and state-owned companies → **our focus**.
  2. Non-traditional entities are persons or legal entities that contract with the state to provide public services or manage public funds.
  3. Political parties or social movements.

Research Design

Overview

1. A field experiment conducted in collaboration with the PGN in the collection of the 2020 ITA index. Allows us to answer:
    - How does the reporting behavior of bureaucrats respond to changes in the salience of oversight?
2. An independent audit of a subset of responses provided by a random sample of public sector (traditional) entities. Allows us to answer:
    - How does the reported data we observe relate to the truth?

Experimental design: treatment level 1

Status quo communication from sector heads in central government about ITA obligations	↔	Status quo communication from sector heads **and** direct communication from PGN watchdog agency via email about ITA obligations

- Interpretation of contrast and ATE:
  - Make PGN's role and use of the data more visible.
  - From ex-post semi-structured interviews with bureaucrats who filled out ITA:
    - PGN is known as a watchdog agency, seen as having some teeth.
    - Entities regularly asked to send data to different central government entities.
    - ... but PGN has distinct oversight powers.

Experimental design: treatment level 1

Status quo communication from sector heads in central government about ITA obligations	← - - →	Status quo communication from sector heads **and** direct communication from PGN watchdog agency via email about ITA obligations

- Interpretation of contrast and ATE:
  - Make PGN's role and use of the data more **visible**.
  - From ex-post semi-structured interviews with bureaucrats who filled out ITA:
    - PGN is known as a watchdog agency, seen as having some teeth.
    - Entities regularly asked to send data to different central government entities.
    - ... but PGN has distinct oversight powers.

Experimental design: treatment level 2

- Nudges—variation in message content—randomly assigned in $2 \times 2 \times 2 \times 2$ factorial design

Status quo communication from sector heads in central government about ITA obligations

Status quo communication from sector heads **and** direct communication from PGN watchdog agency via email about ITA obligations

1. Past oversight: 2019 status

$\times$

2. Future oversight Possible audit

$\times$

3. Reminder email $\approx \uparrow$ dosage

$\times$

4. Training Links to training videos

- Benefits of the "nudge" manipulations:
  - For PGN: sending direct messages has costs, can benefits be maximized at no additional cost?
  - For us: (helps to) unbundle direct communication treatment.

Experimental design: treatment level 2

- Nudges—variation in message content—randomly assigned in $2 \times 2 \times 2 \times 2$ factorial design

Status quo communication from sector heads in central government about ITA obligations

Status quo communication from sector heads **and** direct communication from PGN watchdog agency via email about ITA obligations

1. Past oversight: 2019 status

$\times$

2. Future oversight Possible audit

$\times$

3. Reminder email $\approx \uparrow$ dosage

$\times$

4. Training Links to training videos

- Benefits of the “nudge” manipulations:
  - For PGN: sending direct messages has costs, can benefits be maximized at no additional cost?
  - For us: (helps to) unbundle direct communication treatment.

Experimental design: sample and treatment assignment

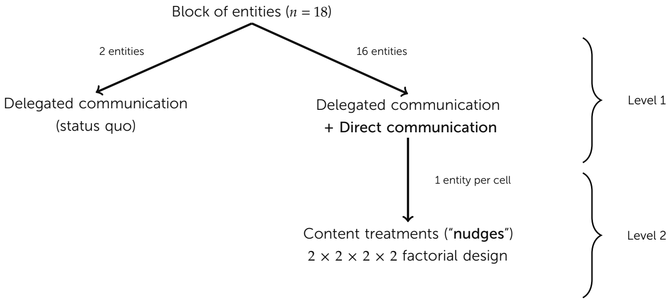
- Sample consists of 6,556 public sector entities in Colombia.
- Blocked assignment: random assignment within blocks of 18 entities:
  - Exact blocking on 2019 ITA completion
  - Minimize Malahnobis distance on (1) PGN's classification of organizational or entity type; (2) department.

Experimental design: sample and treatment assignment

- Sample consists of 6,556 public sector entities in Colombia.
- Blocked assignment: random assignment within blocks of 18 entities:
  - Exact blocking on 2019 ITA completion
  - Minimize Malahnobis distance on (1) PGN's classification of organizational or entity type; (2) department.

Experimental design: sample and treatment assignment

- Sample consists of 6,556 public sector entities in Colombia.
- Blocked assignment: random assignment within blocks of 18 entities:
  - Exact blocking on 2019 ITA completion
  - Minimize Malahnobis distance on (1) PGN's classification of organizational or entity type; (2) department.



Estimation of treatment effects

- We estimate the ATE of direct communication and AMCEs of the nudge treatments effects by OLS:

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

- The estimators of the treatment effects are:
  - β_1: The ATE of direct communication
  - β_2, β_3, β_4, β_5: The AMCEs of the factorial message treatments.

Independent audit design

- Stratified random sample of 2,400 of 6,556 public sector entities
  - Crucially, we sampled entities regardless of whether they **completed the 2020 ITA**.
  - So we observe reporters and non-reporters symmetrically in the audit.
- Provides validation of self-reported transparency practices through an independent audit of a subset of index items:
  - Conducted outside partnership with PGN.
  - Audited items worth 27.75/100 points on the index.
  - We **measure** θ (quality) for this subset of the index.

Independent audit design

- Stratified random sample of 2,400 of 6,556 public sector entities
  - Crucially, we sampled entities regardless of whether they **completed the 2020 ITA**.
  - So we observe reporters and non-reporters symmetrically in the audit.
- Provides validation of self-reported transparency practices through an independent audit of a subset of index items:
  - Conducted outside partnership with PGN.
  - Audited items worth 27.75/100 points on the index.
  - We **measure** θ (quality) for this subset of the index.
- Premise of design: measurement error in measure or θ is independent of measurement error in r.

Independent audit design

- Stratified random sample of 2,400 of 6,556 public sector entities
  - Crucially, we sampled entities regardless of whether they **completed the 2020 ITA**.
  - So we observe reporters and non-reporters symmetrically in the audit.
- Provides validation of self-reported transparency practices through an independent audit of a subset of index items:
  - Conducted outside partnership with PGN.
  - Audited items worth 27.75/100 points on the index.
  - We **measure** θ (quality) for this subset of the index.
- Premise of design: measurement error in measure or θ is independent of measurement error in r.

Measurement of parameters of interest

- **Reported data** (r):
  - Submission of the ITA matrix (yes/no)
  - ITA score (0-100) conditional on submission.
- **Quality** (θ for subset of index):
  - Use index weighting scheme from full index to construct a "true" score between 0 and 27.75 points.
  - We can also reconstruct the reported score on this subset from micro-data.

Measurement of parameters of interest

- Reported data (r):
  - Submission of the ITA matrix (yes/no)
  - ITA score (0-100) conditional on submission.
- Quality (θ for subset of index):
  - Use index weighting scheme from full index to construct a "true" score between 0 and 27.75 points.
  - We can also reconstruct the reported score on this subset from micro-data.
- Mapping to (unobserved) bureaucratic behavior:
  - Recall that $r - \theta = d + \epsilon$ (intentional + unintentional error).
  - By assumption $E[\epsilon] = 0$, so we will examine $E[r - \theta]$.
  - Link between variance of r and effort (e).

Measurement of parameters of interest

- Reported data (r):
  - Submission of the ITA matrix (yes/no)
  - ITA score (0-100) conditional on submission.
- Quality (θ for subset of index):
  - Use index weighting scheme from full index to construct a "true" score between 0 and 27.75 points.
  - We can also reconstruct the reported score on this subset from micro-data.
- Mapping to (unobserved) bureaucratic behavior:
  - Recall that $r - \theta = d + \epsilon$ (intentional + unintentional error).
  - By assumption $E[\epsilon] = 0$, so we will examine $E[r - \theta]$.
  - Link between variance of r and effort (e).

Results I: Experimental Findings

Experimental results

- ATE and AMCE estimates on submission of ITA matrix.

	Completed ITA	
	(1)	(2)
Direct communication	0.029	0.030
	(0.022)	(0.020)
Oversight of past completion	0.000	-0.001
	(0.012)	(0.011)
Possible future audit	-0.005	-0.005
	(0.012)	(0.011)
Direct reminder	0.013	0.012
	(0.012)	(0.011)
Training	-0.008	-0.007
	(0.012)	(0.011)
Observations	6556	6556
Sample	All	All
Block FE		✓
Control mean	0.68	0.68
Control std. dev.	0.47	0.47
DV range	{0,1}	{0,1}

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

Private sector benchmark Past ITA completion

Experimental results

- "Post-treatment" estimand on reported scores.

	Completed ITA		ITA Score	
	(1)	(2)	(3)	(4)
Direct communication	0.029	0.030	-6.066***	-5.886***
	(0.022)	(0.020)	(1.477)	(1.473)
Oversight of past completion	0.000	-0.001	0.587	0.540
	(0.012)	(0.011)	(0.950)	(0.936)
Possible future audit	-0.005	-0.005	-0.453	-0.666
	(0.012)	(0.011)	(0.950)	(0.936)
Direct reminder	0.013	0.012	-2.836***	-2.763***
	(0.012)	(0.011)	(0.949)	(0.932)
Training	-0.008	-0.007	-1.094	-1.191
	(0.012)	(0.011)	(0.950)	(0.939)
Observations	6556	6556	4446	4446
Sample	All	All	Completed	Completed
Block FE		✓		✓
Control mean	0.68	0.68	73.37	73.37
Control std. dev.	0.47	0.47	29.4	29.4
DV range	{0,1}	{0,1}	[0,100]	[0,100]

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

Private sector benchmark

Decomposing the "effect" on scores

- Post-treatment estimates suggests that direct communication →↓ scores, but there are two possible explanations.
    1. **Treatment effect**: Those that would always report submit lower scores when subjected to oversight.
    2. **Selection/compositional change**: Those that report because of treatment report lower average reported scores than those that always report.

- Assuming monotonicity, we can decompose post-treatment estimand into:
    - CATE on always-reporters → Lee (2009) trimming bounds!
    - Average outcomes (scores) of if-treated reporters → boundable if we have post-treatment estimate, treatment effect on reporting, and CATEs.

Decomposition

Decomposing the "effect" on scores

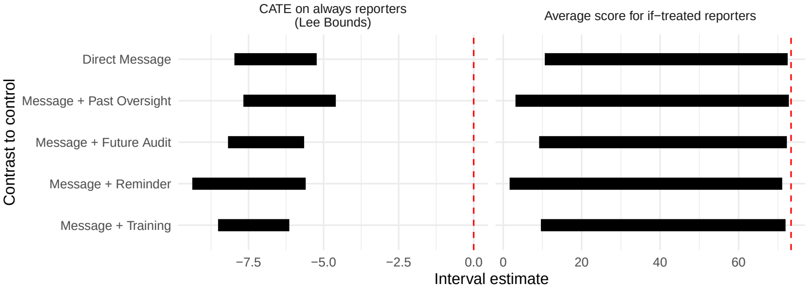
- Post-treatment estimates suggests that direct communication $\rightarrow\downarrow$ scores, but there are two possible explanations.
    1. **Treatment effect**: Those that would always report submit lower scores when subjected to oversight.
    2. **Selection/compositional change**: Those that report because of treatment report lower average reported scores than those that always report.

- Assuming monotonicity, we can **decompose** post-treatment estimand into:
    - **CATE** on always-reporters → Lee (2009) trimming bounds!
    - **Average outcomes** (scores) of if-treated reporters → boundable if we have post-treatment estimate, treatment effect on reporting, and CATEs.

Decomposition

Treatment effects or selection?

- Evidence of both treatment effects and selection:
  - ↓ scores among always reporters.
  - Lower-than-average scores among if-treated reporters.



What do/don't we learn from the experiment?

- Manipulation of the role of the PGN/visibility of oversight:
  - Induce public sector entities to report **less desirable** scores.
  - Change the **composition** of entities that report to the national government.
- What we cannot observe from the experiment:
  - Who **selects** into reporting?
  - **Accuracy**: how do reported scores relate to actual quality?

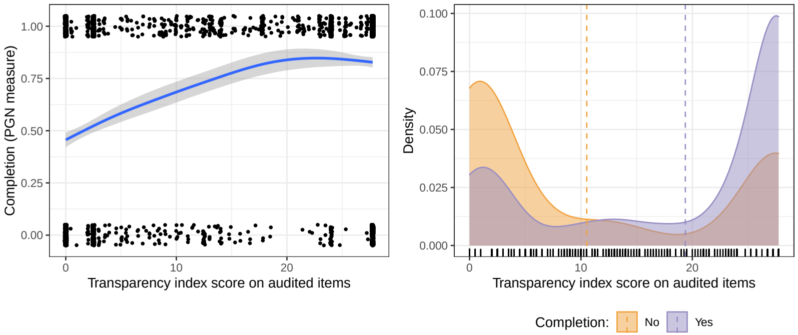
What do/don't we learn from the experiment?

- Manipulation of the role of the PGN/visibility of oversight:
  - Induce public sector entities to report **less desirable** scores.
  - Change the **composition** of entities that report to the national government.
- What we cannot observe from the experiment:
  - Who **selects** into reporting?
  - **Accuracy**: how do reported scores relate to actual quality?

Results II: Audit Findings

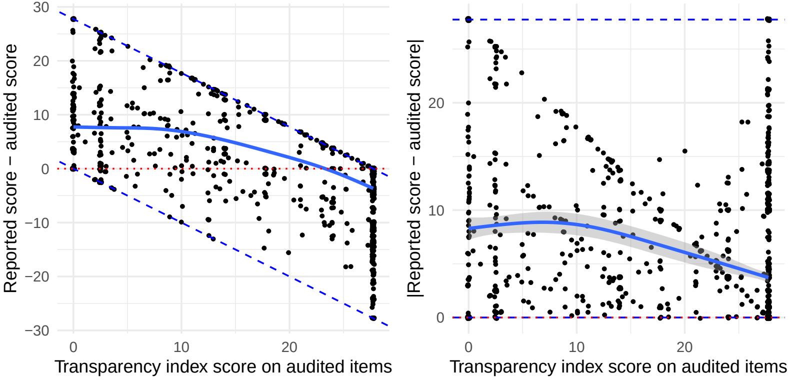
Which entities select into reporting?

- Positive selection into reporting as a function of "true" transparency practices, θ.



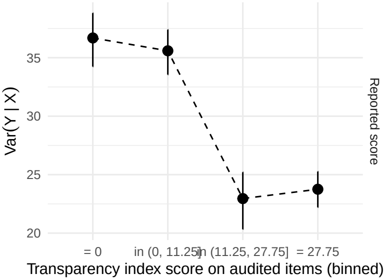
Are reported scores accurate?

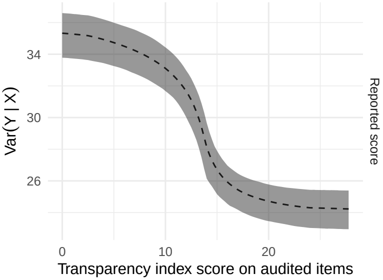
- At lower levels of θ, reported scores are less accurate.
  - In general, scores are over-reported.



When are reported scores less noisy?

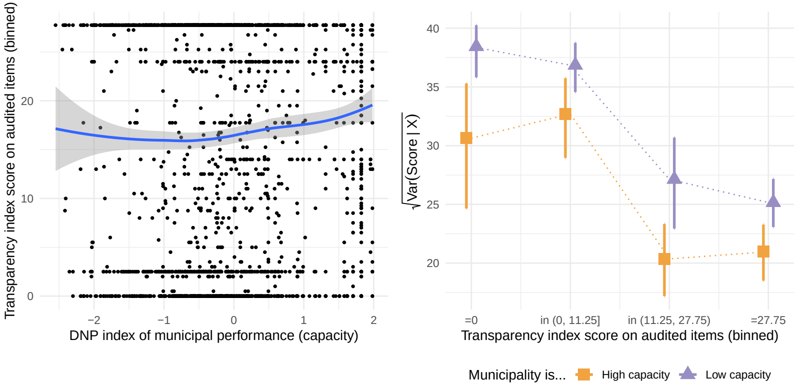
- Variance of reported scores is decreasing in θ.
  - Observed pattern not driven (exclusively) by **state capacity** or **intentional distortion**.
  - Within framework, suggests that lower θ entities exert less effort.





Is this all just variation in state capacity?

- State capacity arguably increases costs of effort, $c(e)$.
  - Should ↓ effort → ↑ variance of ε
- Using Departmento Nacional de Planeación's mendición de desempeño municipal we stratify on **municipal capacity**
  - We still observe decrease in conditional variance in θ.



Discussion

The central government's problem

- By experimenting, we have abstracted away from the central government's problem: When can data collected from decentralized entities be used to inform policies concerning those entities?
  - A preoccupation of central government entities.
- In our problem, recall that there are two policy instruments that central governments set:
  - Probability of oversight: $\rho(r) \in (0,1)$.
  - Penalties imposed when oversight reveals poor outcomes: $P(\theta, r) > 0$.

The central government's problem

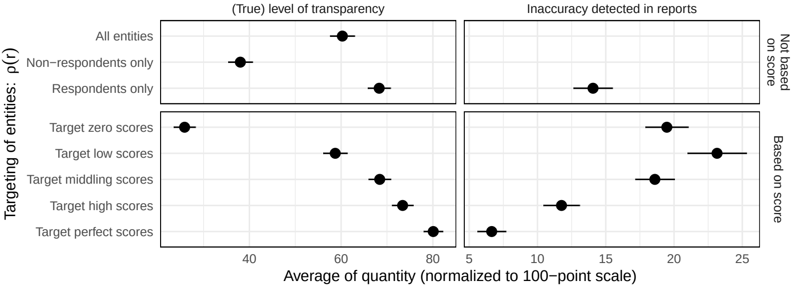
- By experimenting, we have abstracted away from the central government's problem: When can data collected from decentralized entities be used to inform policies concerning those entities?
  - A preoccupation of central government entities.
- In our problem, recall that there are two policy instruments that central governments set:
  - Probability of oversight: $\rho(r) \in (0,1)$.
  - Penalties imposed when oversight reveals poor outcomes: $P(\theta, r) > 0$.
- Can our data tell us about what the central government should do?
  - Not exactly: changes in these parameters (if known) should change reporting behavior.
  - But, we can think about what the government would "see" under different instruments fixing the "mechanism" from the experiment.

The central government's problem

- By experimenting, we have abstracted away from the central government's problem: When can data collected from decentralized entities be used to inform policies concerning those entities?
    - A preoccupation of central government entities.
- In our problem, recall that there are two policy instruments that central governments set:
    - Probability of oversight: $\rho(r) \in (0,1)$.
    - Penalties imposed when oversight reveals poor outcomes: $P(\theta, r) > 0$.
- Can our data tell us about what the central government should do?
    - Not exactly: changes in these parameters (if known) should **change reporting behavior**.
    - But, we can think about what the government would "see" under different instruments fixing the "mechanism" from the experiment.

Reporting behavior conditions what PGN find

- We **simulate** what the PGN should observe depending on how they target preventative efforts, $\rho(r)$.
  - PGN sees **reports** (r).
  - We observe **reports** (r) and the **truth** (⊠).



Take-aways

- Administrative data as a bureaucratic output and political outcome matters:
  - to you as a producer/consumer of empirical social science.
  - to governments that collect data in order to use it.
- Data production as an understudied aspect of intergovernmental relations.

Take-aways

- Administrative data as a bureaucratic output and political outcome matters:
  - to you as a producer/consumer of empirical social science.
  - to governments that collect data in order to use it.
- Data production as an understudied aspect of intergovernmental relations.
- For policy: limits to data-driven governance as a efficiency-enhancing reform.

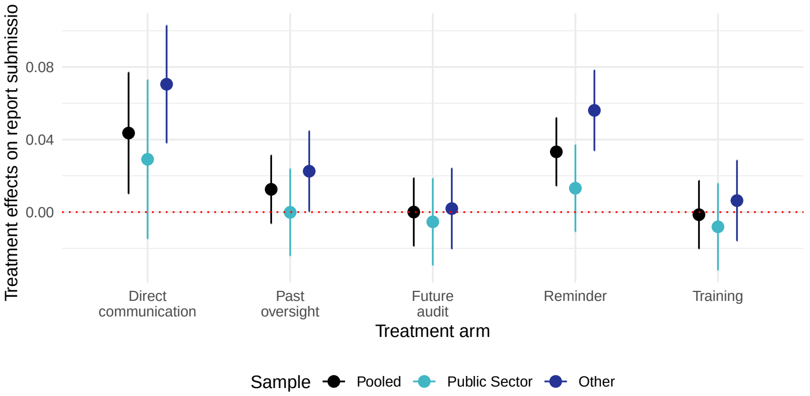
Take-aways

- Administrative data as a bureaucratic output and political outcome matters:
    - to you as a producer/consumer of empirical social science.
    - to governments that collect data in order to use it.
- Data production as an understudied aspect of intergovernmental relations.
- For policy: limits to data-driven governance as a efficiency-enhancing reform.

Thank you!

Benchmark: Private Sector Entity Completion

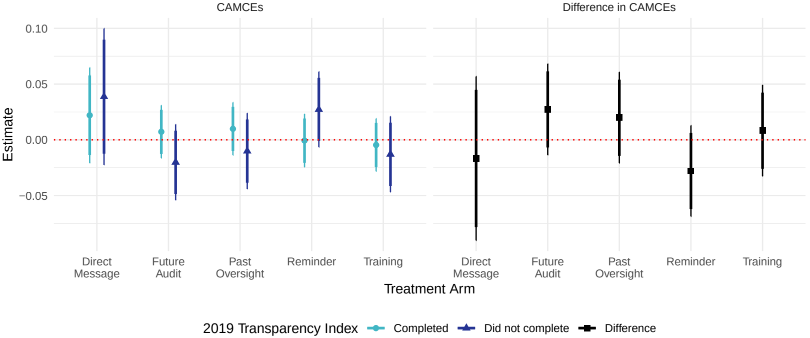
- We estimate the ATE of direct communication and the AMCEs of the messages in the factorial design. Estimator
- Outcome: Indicator for ITA data submission.



Back

Was treatment solely informational?

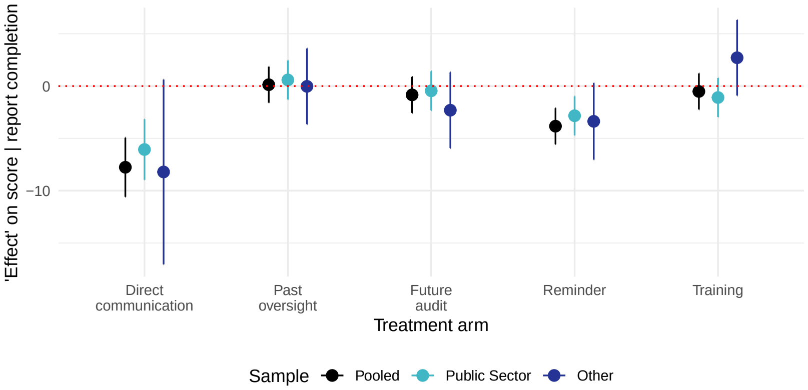
- We estimate the CATE of direct communication and the CAMCEs of the messages conditioning on 2019 ITA completion. Estimator
- Outcome: Indicator for ITA data submission.



Back

Benchmark: Private Sector Entity Scores

- We estimate the same specification with scores as the outcome, but condition on reporting.
- Estimates of a post-treatment estimand that contains both a causal effect and a "selection" effect.



Back

Decomposition of post-treatment estimand

- Under monotonicity, there are three strata: Always reporters, if-Treated reporters, Never reporters
  - Denote types by $j \in \{A, T, N\}$ with shares π_j, where $\sum_j \pi_j = 1$.
- The "post-treatment" estimand is $\mathscr{P}$. We can express this quantity as:

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

- We have point estimates for: $\mathscr{P}$, $\frac{\pi_A}{\pi_A+\pi_T}$, $\frac{\pi_T}{\pi_A+\pi_T}$, and $E[S(Z=0)|j=A]$ (under monotonicity).
- We can use Lee (2019) bounds to generate an interval estimate of $CATE$.
- We can then use algebra to generate an interval estimate of $E[S(Z=1)|j=T]$.

Back