Bureaucratic Incentives and the Production of Administrative Data

Tara Slough

New York University

“Good data” and “bad data”

Social scientists often say a setting has “good data” or “bad data.”

Typically a statement about administrative data quality or availability.

Variation often attributed to (state) capacity:

- Lack of capacity $\longrightarrow$ “bad data”/no data.
- … but most administrative data is collected to be *used*. Shapes:
  - What data is collected
  - How data is collected
  - What gets reported $\longrightarrow$ data quality

Decentralized data collection

A process through which data is solicited by the center and reported by peripheral entities

Examples:

- National gov't soliciting information from state/local gov'ts
- International organization soliciting information from member states
- Multinationals requesting data from local subsidiaries

In context of national administrative data $\longrightarrow$ bureaucrats central to many forms of data collection.

Decentralized data collection: example #1

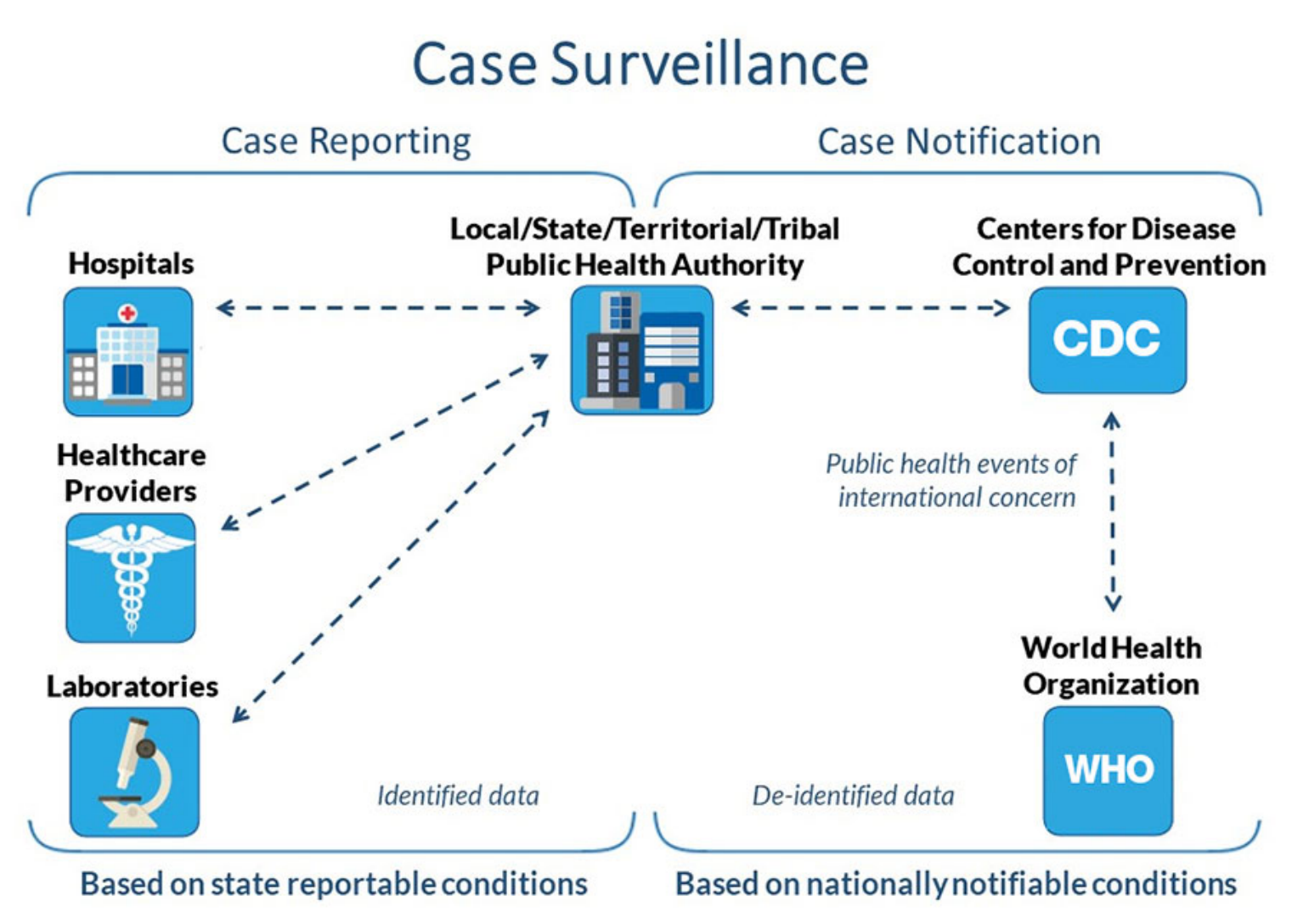


Figure: Source: United States Centers for Disease Control and Prevention.

Decentralized data collection: example #2

Public Law 104–13
104th Congress

An Act

To further the goals of the Paperwork Reduction Act to have Federal agencies become more responsible and publicly accountable for reducing the burden of Federal paperwork on the public, and for other purposes.

May 22, 1995
[S. 244]

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

Paperwork Reduction Act of 1995.
Information resources management.
44 USC 101 note.

SECTION 1. SHORT TITLE.

This Act may be cited as the "Paperwork Reduction Act of 1995".

SEC. 2. COORDINATION OF FEDERAL INFORMATION POLICY.

Chapter 35 of title 44, United States Code, is amended to read as follows:

Decentralized data collection: example #2

Public Law 104–13
104th Congress

An Act

To further the goals of the Paperwork Reduction Act to have Federal agencies become more responsible and publicly accountable for reducing the burden of Federal paperwork on the public, and for other purposes.

May 22, 1995
[S. 244]

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

Paperwork Reduction Act of 1995.
Information resources management.
44 USC 101 note.

SECTION 1. SHORT TITLE.

This Act may be cited as the "Paperwork Reduction Act of 1995".

SEC. 2. COORDINATION OF FEDERAL INFORMATION POLICY.

Chapter 35 of title 44, United States Code, is amended to read as follows:

"§ 3501. Purposes

"The purposes of this chapter are to—

"(1) minimize the paperwork burden for individuals, small businesses, educational and nonprofit institutions, Federal contractors, State, local and tribal governments, and other persons resulting from the collection of information by or for the Federal Government;

"(2) ensure the greatest possible public benefit from and maximize the utility of information created, collected, main-

This paper: Bureaucratic incentives $\longrightarrow$ data quality

Framework: Decentralized data collection as a set of related agency problems

- Between politicians and bureaucrats within
  i.) Central government
  ii.) **Local governments** $\longleftarrow$ This paper
- Between national and local governments

This paper: Bureaucratic incentives $\longrightarrow$ data quality

Framework: Decentralized data collection as a set of related agency problems

Empirics: Link features of bureaucratic incentives and data outputs

- Treatments: **Original survey** of data-producing bureaucrats in Colombia
- Outcomes: **Administrative microdata** on contracts and welfare eligibility

This paper: Bureaucratic incentives $\longrightarrow$ data quality

Framework: Decentralized data collection as a set of related agency problems

Empirics: Link features of bureaucratic employment to data outputs.

Results: Bureaucratic turnover, independence from local politician leaves detectable footprint in administrative microdata.

- Contract data: rate, quality of entries (both on average and variance).
- Social welfare elibility data: share of poor (beneficiary) households; rate of growth of database.

This paper: Bureaucratic incentives $\longrightarrow$ data quality

Framework: Decentralized data collection as a set of related agency problems

Empirics: Link features of bureaucratic employment to data outputs

Results: Bureaucratic turnover, independence from local politician leaves detectable footprint in administrative microdata.

Implications: Use of data by government and third parties (e.g., scholars).

- Social welfare data (planned): Changes in funds allocated to households across municipalities.
- Contract data (planned): Benchmark distortions to documented effects on contracting

Related literature

1. Quality of state data e.g., Jerven (2013), Edmond (2013), Lorentzen (2014) Wallace (2016), Kerner et al. (2017), Guriev and Treisman (2019), Martínez (2021), Brambor et al. (2020), Angrist et al. (2021)

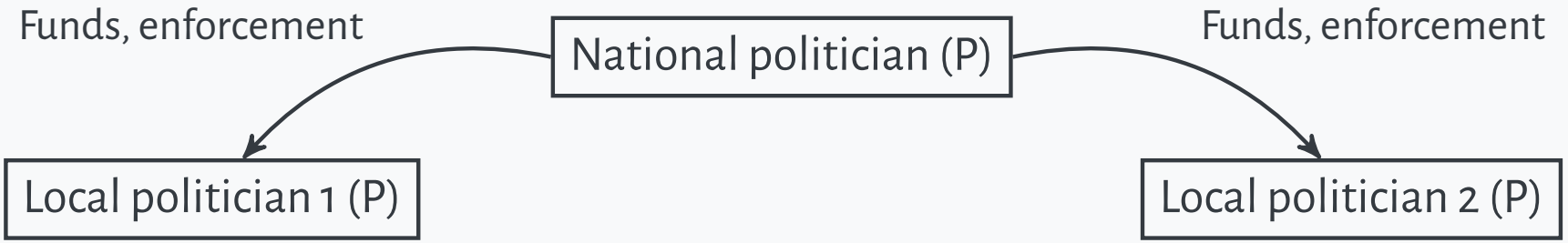
2. Individual-level data production as an interaction between states and citizens Lee and Zhang (2017); Bowles (2024); Sánchez-Talanquer (2020)

3. Data production by bureaucrats/bureaucratic agencies Cook and Fortunato (2022); Eckhouse (2022); Garbiras-Díaz and Slough (2023)

Theoretical Framework

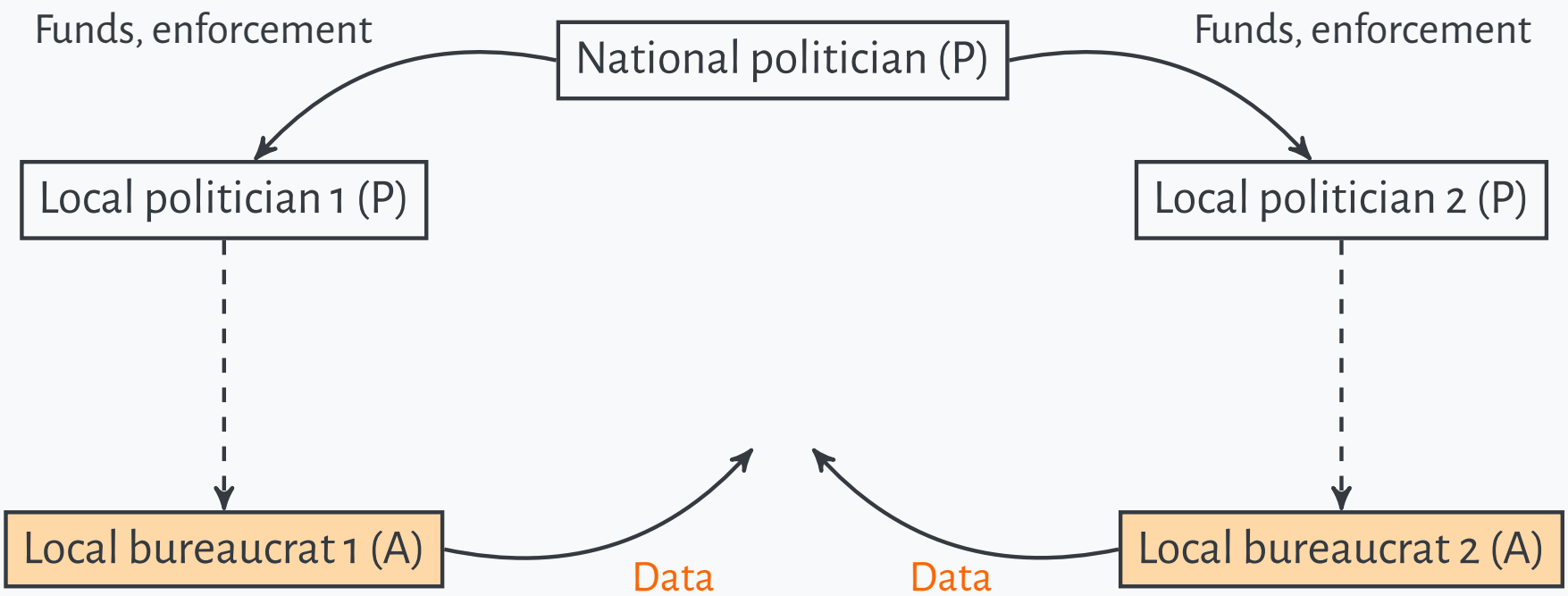
The setting

Distribution of resources or enforcement from center to localities.



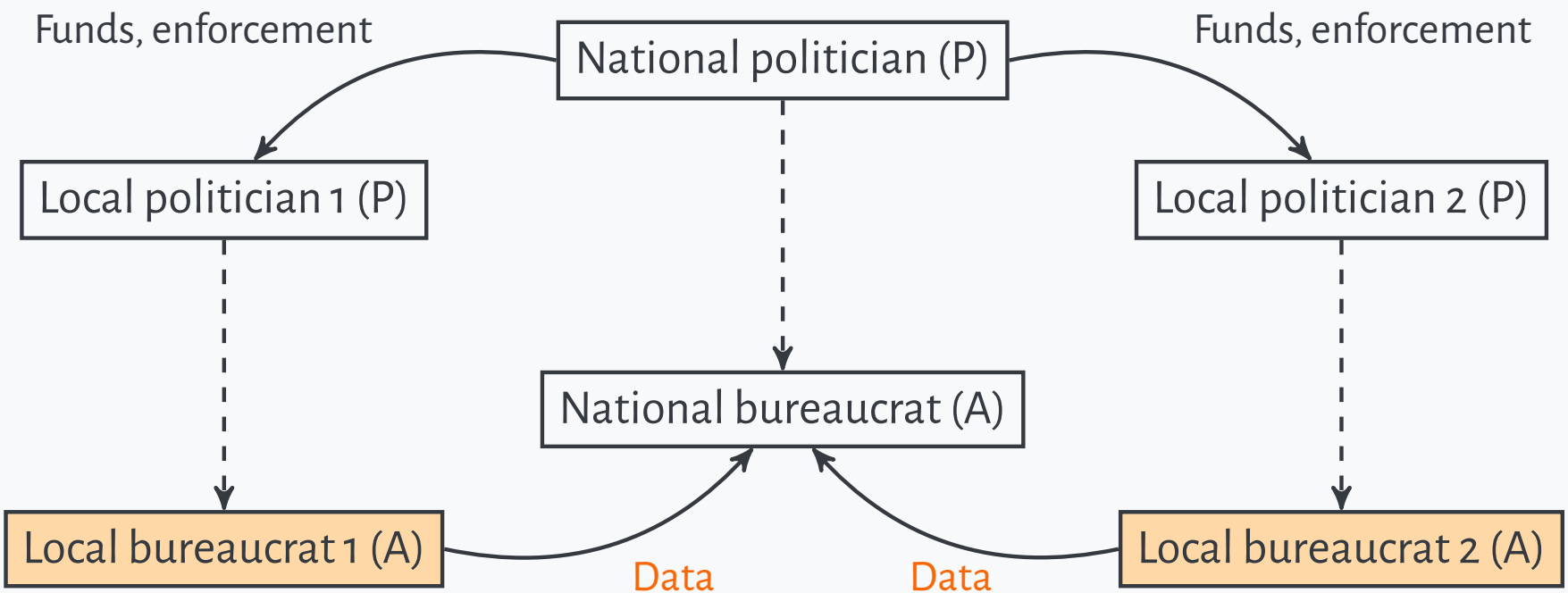
The setting

Distribution of resources or enforcement from center to localities.



The setting

Distribution of resources or enforcement from center to localities.



Conflict over data production

Suppose that the national government uses data to allocate transfer to localities.

National politician $\longrightarrow$ accuracy-motivated:

- Not to be confused with welfare-oriented or technocratic.
- Distributive policy designed by politician: political goals already "baked in"

Local politician $\longrightarrow$ seeks to maximize transfers from center:

- Maximize resources available (transfers) or minimize cumbersome compliance/penalties (enforcement)
- Alternatively: maximize transfers to citizens in locality

. . . **but** we observe variation in the reported data.

Constraints on local politicians

Use of data is concealed by national government: generates uncertainty about optimal report (Camacho and Conover, 2014)

Central government engages in oversight to deter and/or punish misrepresentation (Cook and Fortunato, 2022; Garbiras-Díaz and Slough, 2024.)

Agency problems: Local politicians must rely on bureaucrats to gather, submit data:

- Bureaucrats may not share politician's **preference** for maximizing transfers
- Bureaucrat's **effort is costly**, unobserved

Bureaucratic selection and incentives

When might a bureaucrat's report come closer to a politician's ideal?

- Politician selects a bureaucrat with preferences closer to her own.
- Politician incentivize effort via monitoring, imposition of penalties.

Limitations to this strategy:

- Civil service systems constrain (to varying degrees)
- Regular churn limits expertise accumulation
- Training, consultation with central government bureaucrats

All else equal:

- Bureaucratic turnover $\longrightarrow$ changes in data outputs, potentially closer to politician's ideal.
- Stronger incentives $\longrightarrow$ production of reports closer to politician's ideal.

Case Context

Context: Colombia

Most populous unitary state in the Americas.

Decentralization reforms transferred service provision to local governments

- Local governments rely on center for funds, programs
- National government relies on info. reported by local governments

This project: two widely-used “big” datasets reported by local bureaucrats:

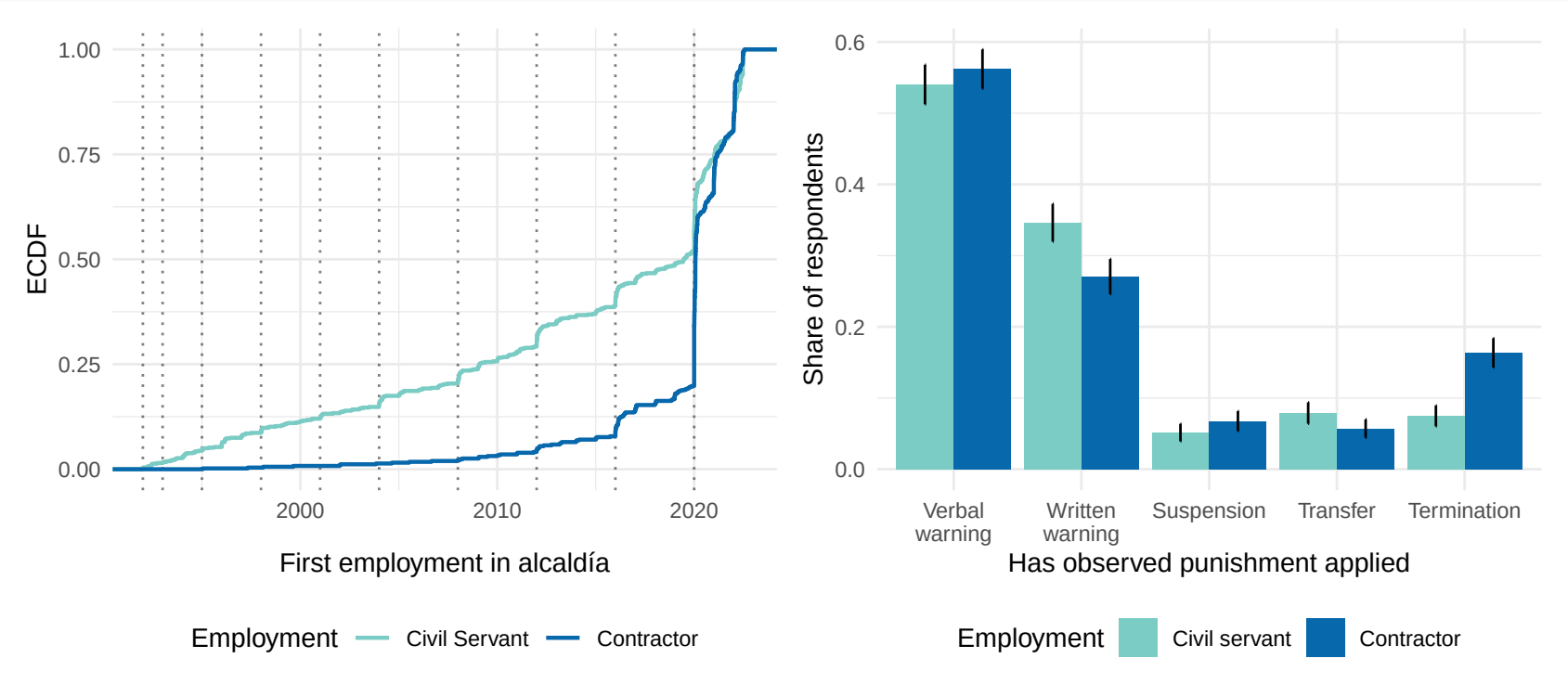
- Procurement databases (SECOP-I and SECOP-II): used in enforcement
- Means-testing surveys (SISBÉN IV): used to target benefits

Local bureaucrats

High variation in recruitment, professionalism, and outputs

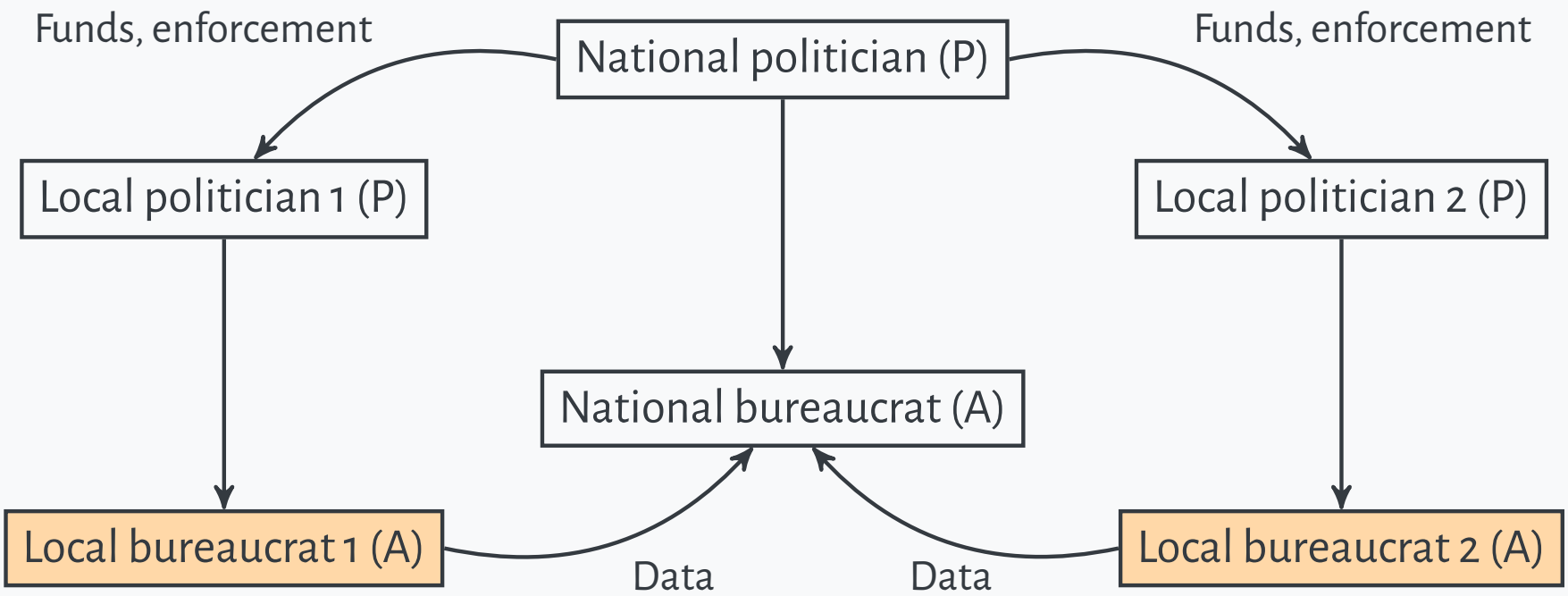
Important source of variation in incentives: civil servant vs. contractor

- Contractors (re-)appointed by mayor; serve shorter tenures
- Subjected to harsher penalties, more oversight than civil servants



Research Design

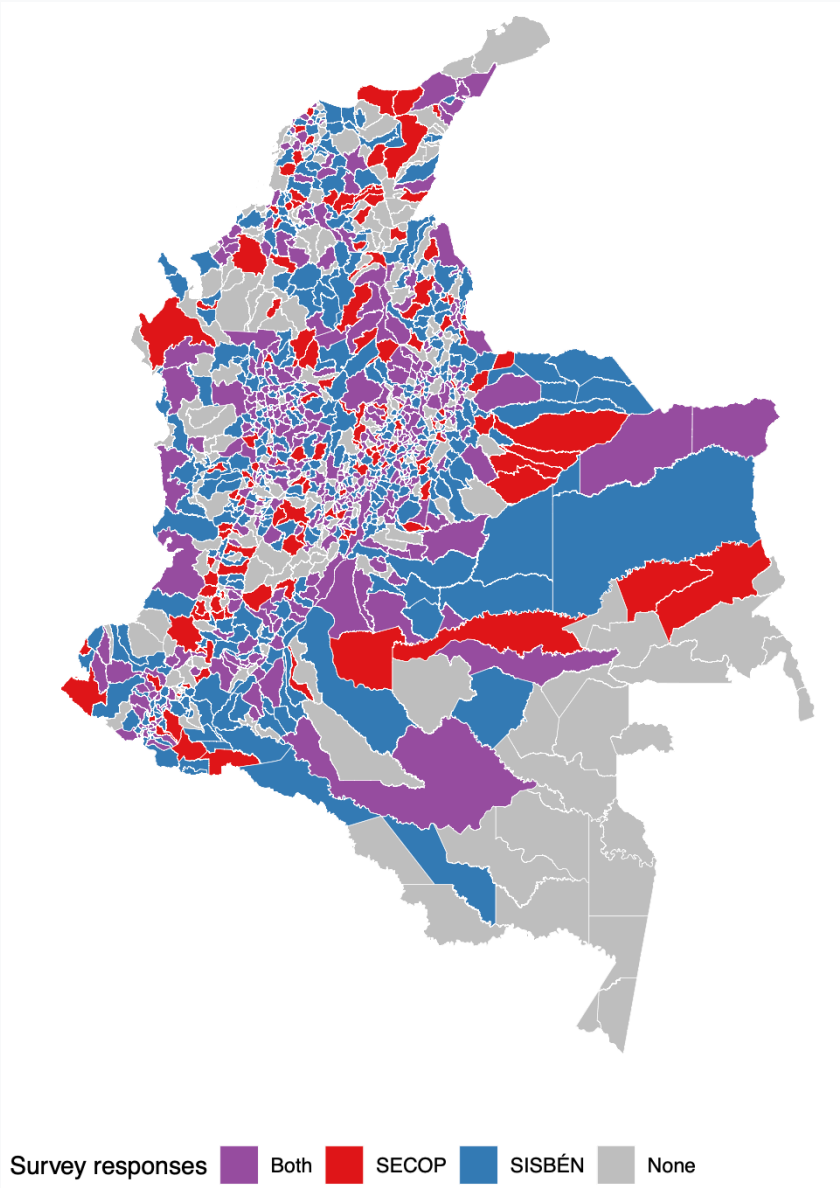
Research Design I: Survey



Original survey of local bureaucrats who submit relevant admin. data.

- Identified relevant bureaucrats from 1,026 and 1,074/1,102 municipalities for SECOP and SISBÉN, respectively.
- Survey administered by email and phone from May-July 2022
- Measure attributes of bureaucrats, perceived oversight in local government, from central government

Survey Response Rate/Sample



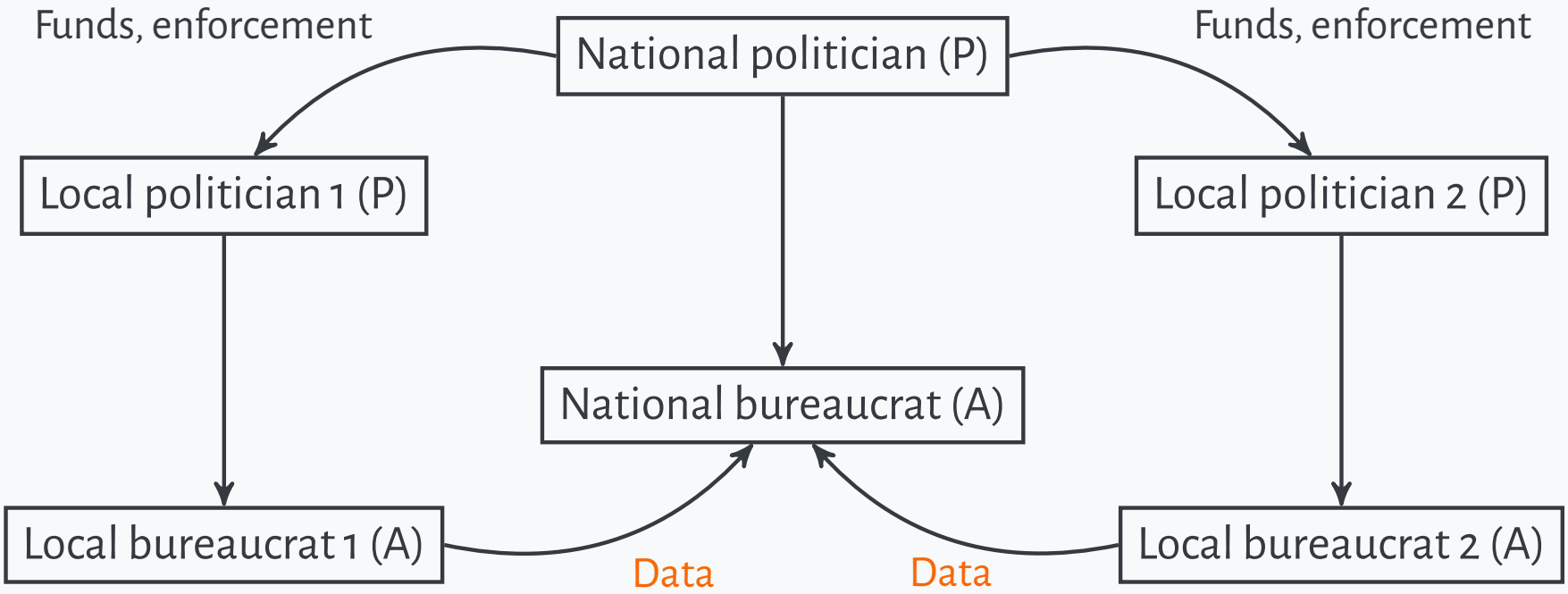
Response rates:

- SECOP: 496/1,026 (48.3%)
- SISBÉN: 752/1,074 (70.0%)

Survey samples resemble population of all municipalities in terms of:

- Population/census covariates
- Geography (regions)
- Administrative classifications, DNP governance indices

Research Design II: Administrative Data



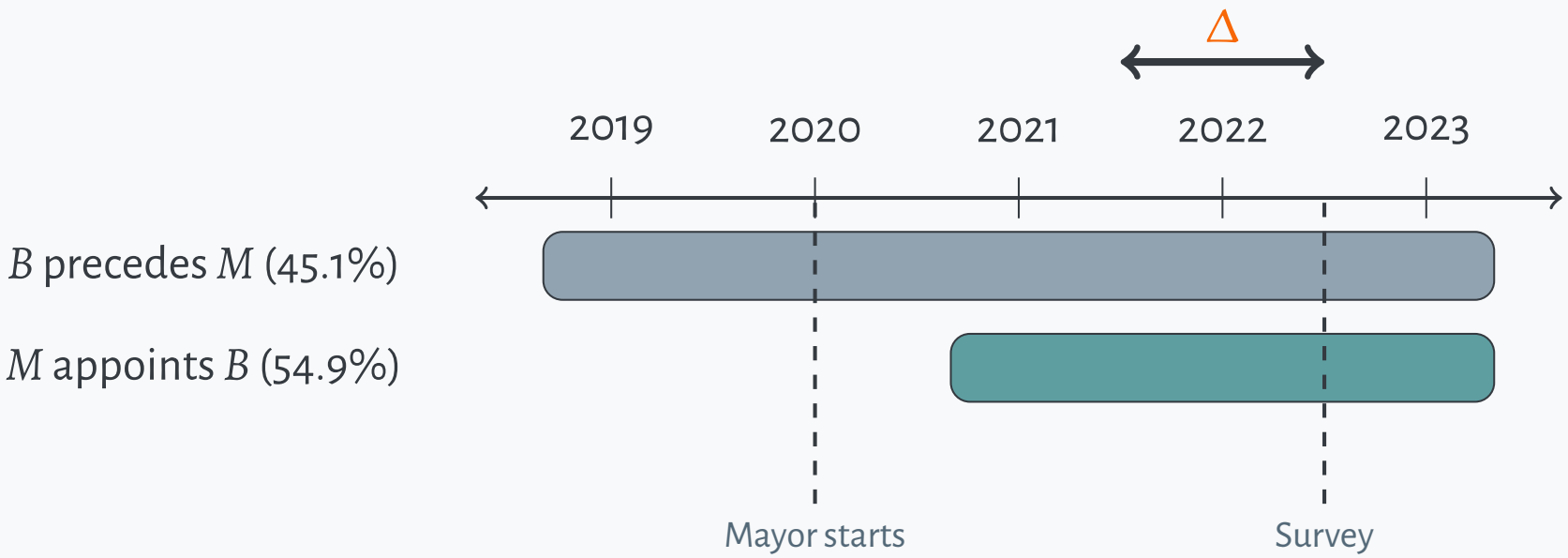
Measure reporting behavior through resultant administrative data

- Measures of *effort* and *data quality*

Administrative data #1: Social program beneficiaries

SISBÉN IV rolled out (nationally) in 2021.

- Observe random cross-sections in mid-2021, 2022, 2023



Examine how bureaucratic tenure and contractor status affect Δ in:

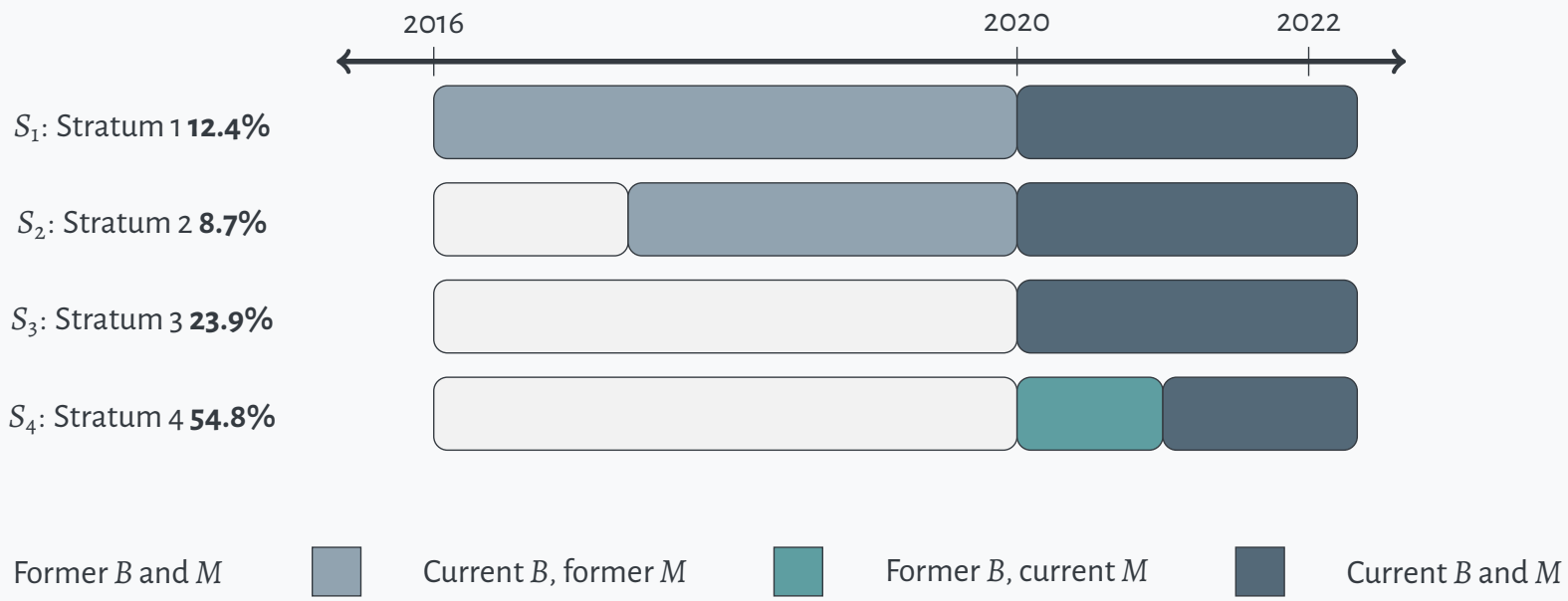
- Share of households in (various states of) poverty ($\longrightarrow$ beneficiaries)
- Total households registered (effort)

Estimation strategy | Sample municipal classification

Administrative data #2: Contracts data

SECOP-I adopted in >95% of municipalities by 2016; used widely until 2022.

- Contracts aggregated to municipality-month level.



Examine how changes in bureaucrat and mayor affect:

- Number, type, value of reported contracts
- Length of information provided about contract (effort)

Estimation strategy | Use of SECOP

Results

Describing bureaucratic incentives

Limited correlation in the duration, type, and reported characteristics of public employment within the same local government.

- Staffing, oversight strategies are domain-specific.
- Implication for data: extent of strategic misreporting by same locality should vary across measures.

	Full survey sample		Both surveys complete		Within municipality
Variable	SECOP	SISBÉN	SECOP	SISBÉN	Correlation (ρ)
Short-tenured (appointed since 2020)	0.67	0.55	0.70	0.56	-0.06
Contractor	0.55	0.34	0.55	0.34	0.15
Mayor supervises	0.39	0.37	0.39	0.39	0.05
Civil servant supervises	0.74	0.90	0.72	0.91	0.06
Supervisor sets goals	0.45	0.70	0.45	0.69	0.03
Oversight frequency (0-5 scale)	3.69	3.13	3.63	3.17	0.06
Perceived freedom (1-5 scale)	3.22	3.67	3.21	3.74	-0.06
Any communication with central government	0.23	0.51	0.22	0.51	0.04

Table: Measures of bureaucrat incentives within the *alcaldía*. All variables are binary unless otherwise noted.

Disaggregating the treatments:

Type of contract predicts tenure length (and vice versa), but differences perceptions of other management practices is limited.

- Implication: Treatments are compound; but perceptions of oversight do not vary substantially by bureaucrat tenure or contract type.

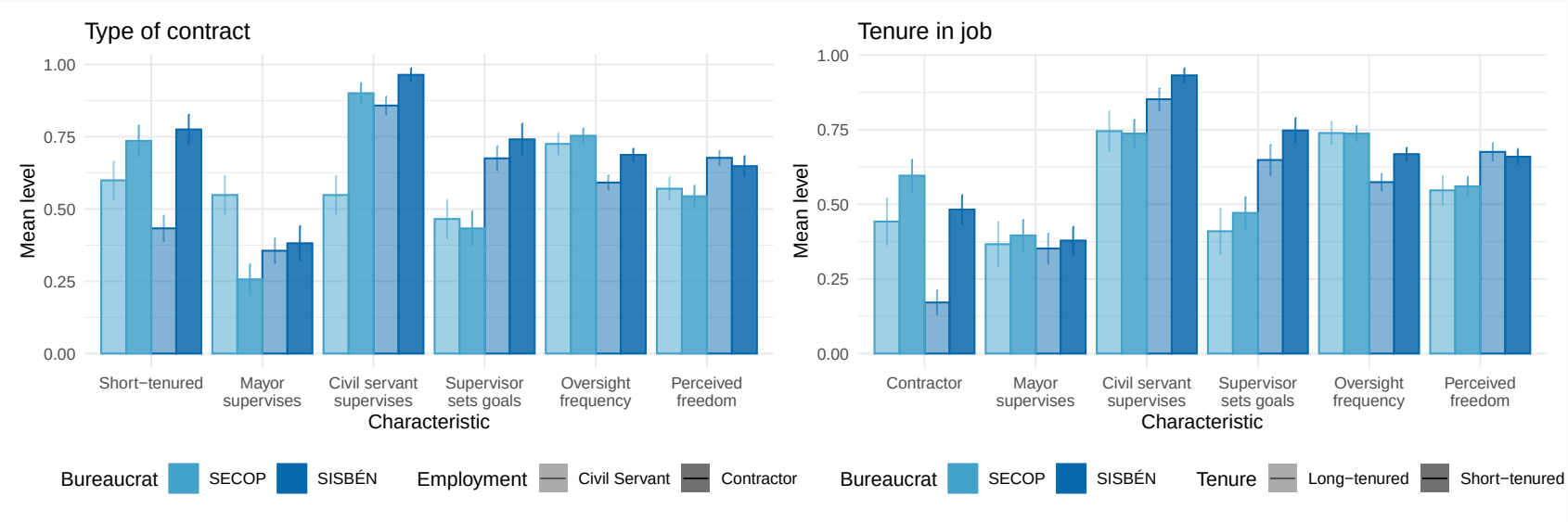


Figure: Rates of other personnel selection and management practices reported by civil servants vs. contractors (left) and long- and short-tenured bureaucrats (right).

SISBÉN: Short-tenured bureaucrats, contractors $\longrightarrow \uparrow$ poverty

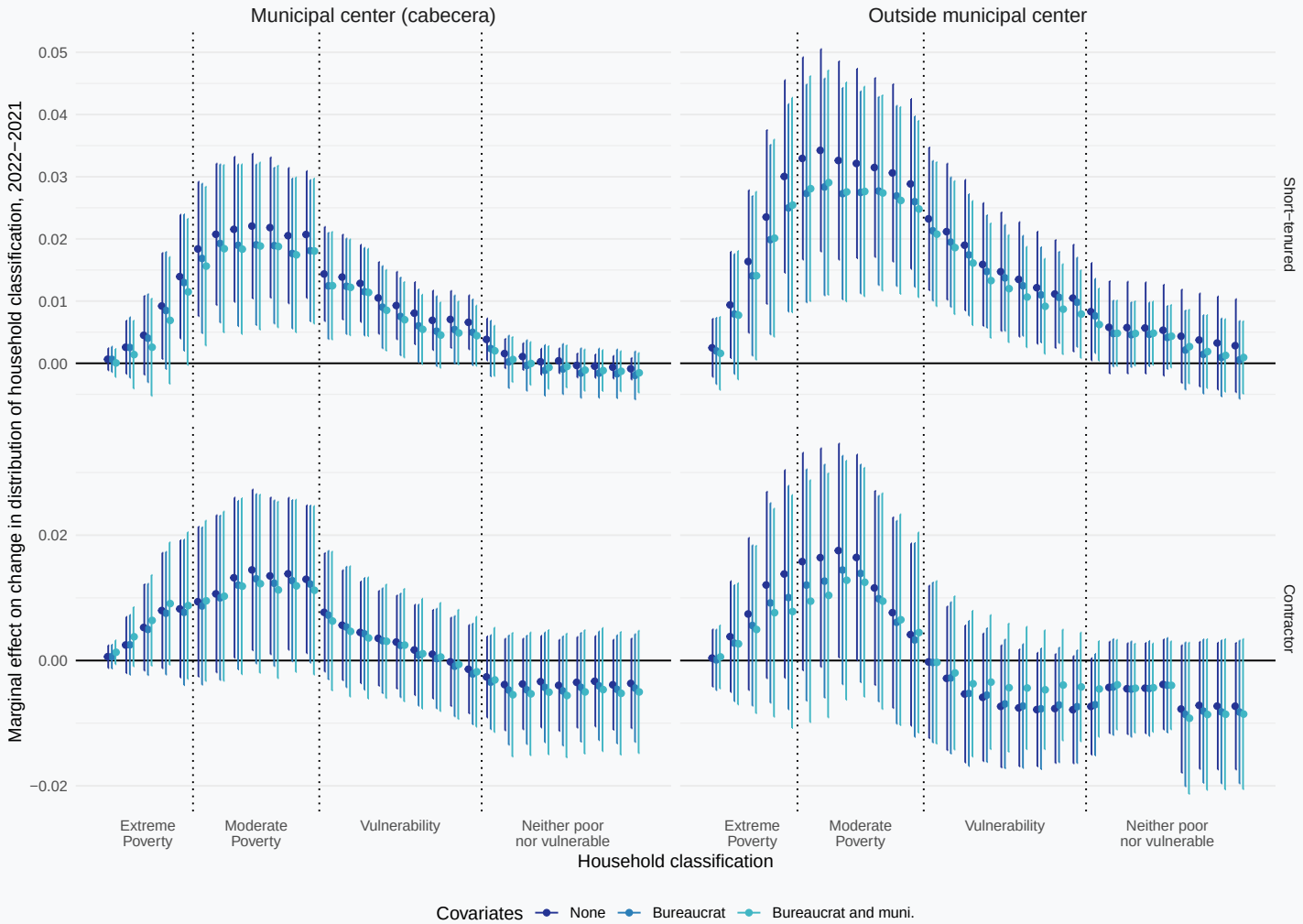


Figure: Changes in cumulative mass function of household classification across 30 ranked categories.

SISBÉN: Interpretation

These shifts are *not* driven by:

- Differences in baseline (2021) enrollment per capita
- Differences in rate of new enrollments between 2021 and 2022

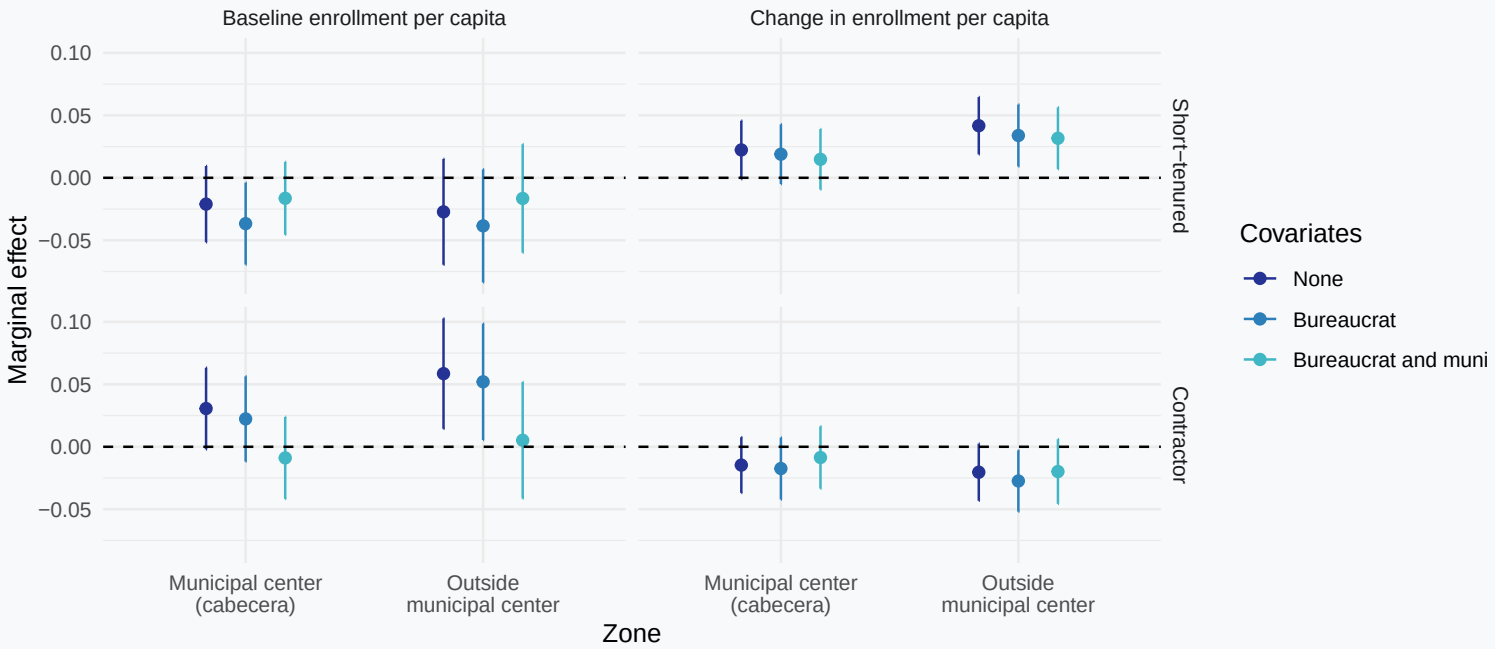


Figure: Marginal effects of bureaucratic employment on baseline enrollment (left) and changes in enrollment per capita (right).

SISBÉN: Interpretation

Instead, shifts toward greater poverty are driven by some combination of:

- Reclassification of surveyed households toward poorer classifications:
  - Non-poor categories more likely to *shrink* with short-tenured or contractor bureaucrat
  - … despite 15% growth in SISBÉN registrations
- New households are more likely to be poor.
  - Growth greater (17% vs. 14%) outside the municipal center $\longrightarrow$ poorer.

Note: From 2021-2022, Colombia's economy grew by 11%. No detectable change in inequality.

SECOP: Changes in contracts data

Comparing effects estimated within three distinct subsamples:

- Turnover reduces reported # of contracts, non-competitive contracts.
- Change in mayor (with or without bureaucrat) $\longrightarrow$ changes in reporting system, effort devoted to reporting $\longrightarrow$ caution warranted.

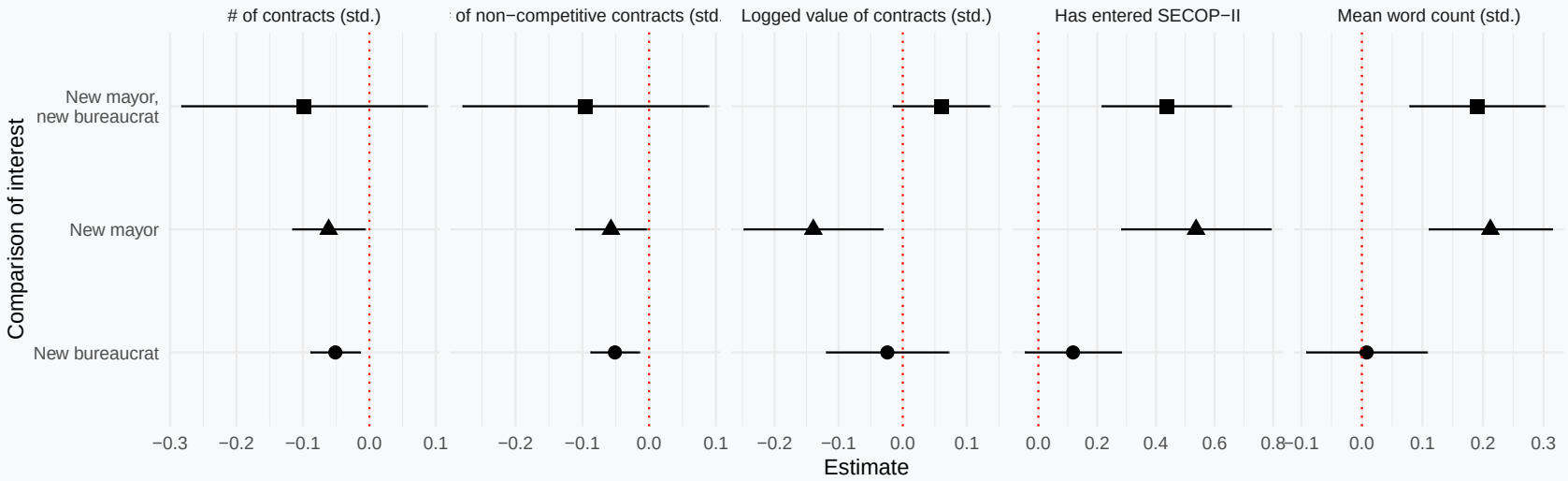


Figure: Effects (ATTs) of changes in mayor, bureaucrat, or both on reporting in SECOP-I. All outcomes are standardized and interpretable as pre-treatment standard deviations.

SECOP: Effects on variance

Turnover of bureaucrat leads to noisier-than-expected reporting behavior.

- Outcome: difference from outcomes predicted from 2016-2019 behavior.
- Permutation-based approach to guard against Type-I errors.

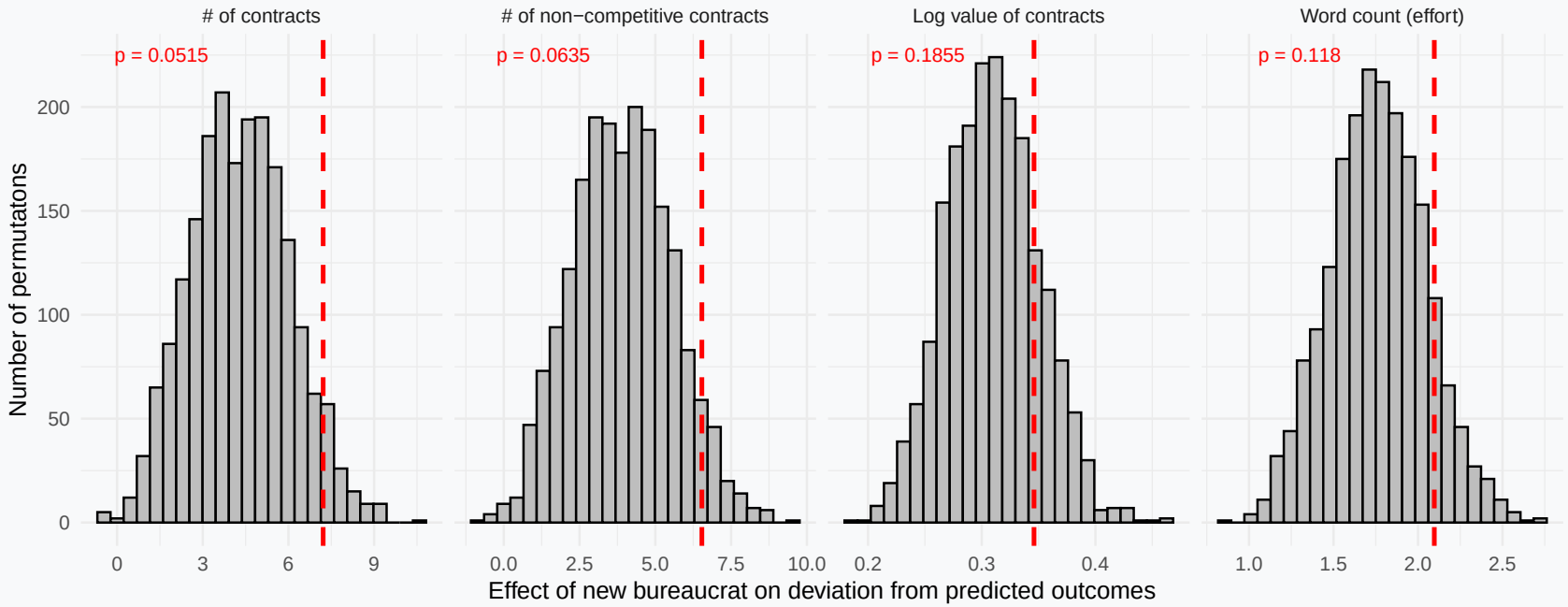


Figure: Effects (ATTs) of changes in bureaucrat alone on changes from predicted reportisin SECOP-I. All outcomes are standardized and interpretable as pre-treatment standard deviations.

SECOP: Interpretation

Effects of bureaucratic, mayoral turnover on contract data is *smaller in magnitude* and *less consistent* than changes in SISBÉN.

- Time scale is important: if differences cumulate over time, number of "missing contracts" per year is higher than per month.
- Potentially fewer incentives for distortion for politicians:
  - Use in enforcement is not as obvious, known to politicians.
  - National government has capacity for auditing, punishing mayors.

Implications and Discussion

Implications: Next Steps

Small changes in state data $\longrightarrow$ non-trivial consequences for:

- Who gets what
- What can be learned by third-party data users

Domain in which small actions by bureaucrats can have large impact.

Quantify consequences of bureaucratic employment-induced distortions on:

- SISBÉN: Amount of transfers to households in municipality
- SECOP: Inferences made by academics using contracts data (e.g., Ruiz (2021), Rueda and Ruiz (2022), Gulzar, Rueda, and Ruiz (2022), Harding et al. (2023), and Carreri and Martinez (2023))

Discussion

Information or data production a central task of the state

Institutional design that facilitates data production is consequential:

- Role of bureaucrats is understudied
- Data production as an important forum for intergovernmental interactions
- Common institutional design across contexts

Important to move beyond state capacity as cause of, proxy for, and consequence of state information

- For policy: improving the quality of state data

Thank you!

Comments welcome!
tara.slough@nyu.edu

SISBÉN Estimation Strategy

I estimate the following by OLS:

$$Y^{(c)}_{mz,t} - Y^{(c)}_{mz,t-1} = \gamma_B \text{Short tenure}_m + \gamma_C \text{Contractor}_m + \boldsymbol{\kappa}\mathbf{X}_m + \epsilon_{mz} \qquad (1)$$

where:

- m indexes municipalities
- z indexes zones of municipalities (*cabecera* versus non-*cabecera*)
- t indexes years
- (c) indexes a specific category/classification of households
- $\mathbf{X}_m$ is a vector of bureaucrat and municipality covariates

Standard errors are clustered at the level of m.

Research design

SECOP Estimation Strategy

The following are OLS regression models. Graphs report β_M, β_{MB}, and β_B:

- For $m \in S_1 \cup S_2$, Cur. bureau.$_{mt} = 1$:

$$Y_{mtc} = \beta_M \text{Current mayor}_t + \boldsymbol{\lambda}_m + \boldsymbol{\gamma}_c + \epsilon_{mt} \tag{2}$$

- For $m \in S_3$:

$$Y_{mtc} = \beta_{MB} \text{Current mayor}_t + \boldsymbol{\lambda}_m + \boldsymbol{\gamma}_c + \epsilon_{mt} \tag{3}$$

- For $m \in S_4$, Current mayor$_t = 1$:

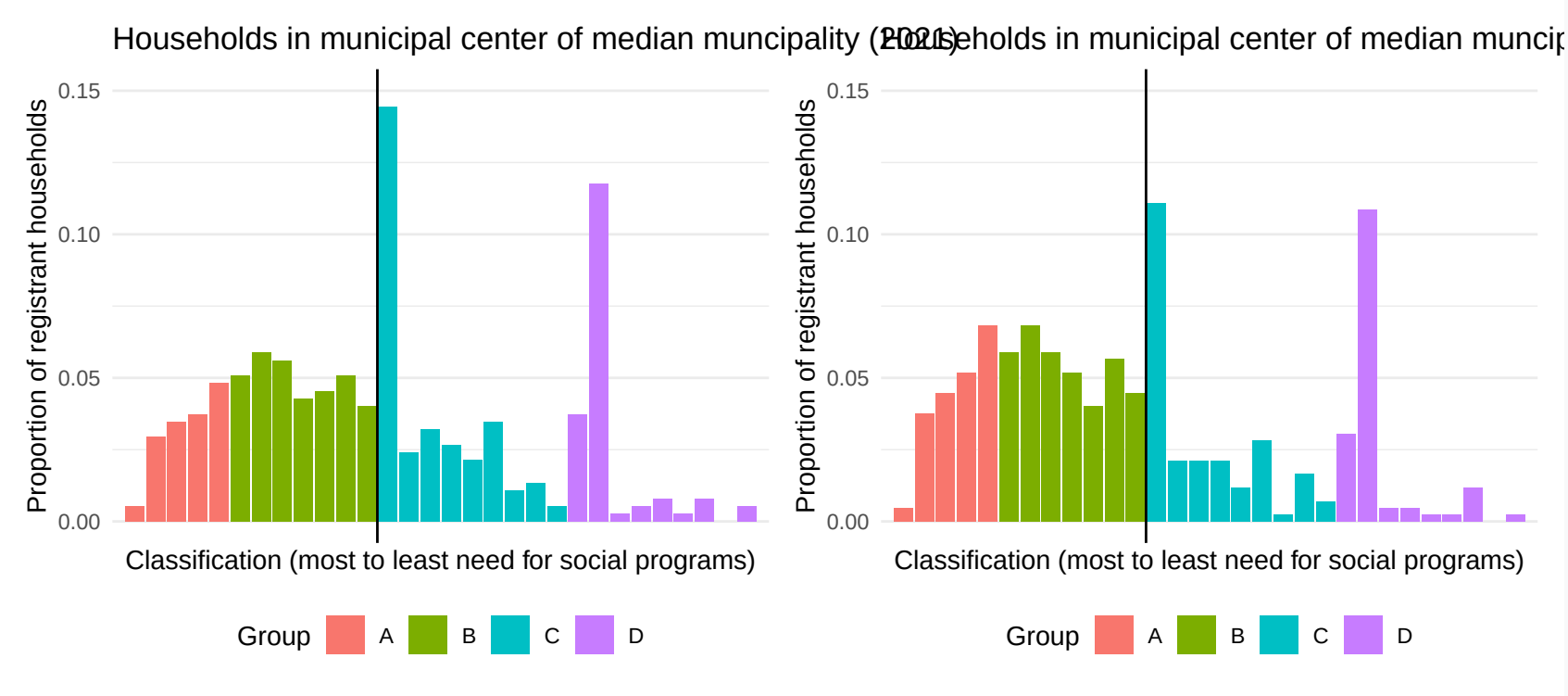
$$Y_{mtc} = \beta_B \text{Current bureaucrat}_{mt} + \boldsymbol{\lambda}_m + \boldsymbol{\gamma}_t + \epsilon_{mt} \tag{4}$$

Standard erros are clustered at the municipality (m) level.

Outcome construction to measure effects on variance:

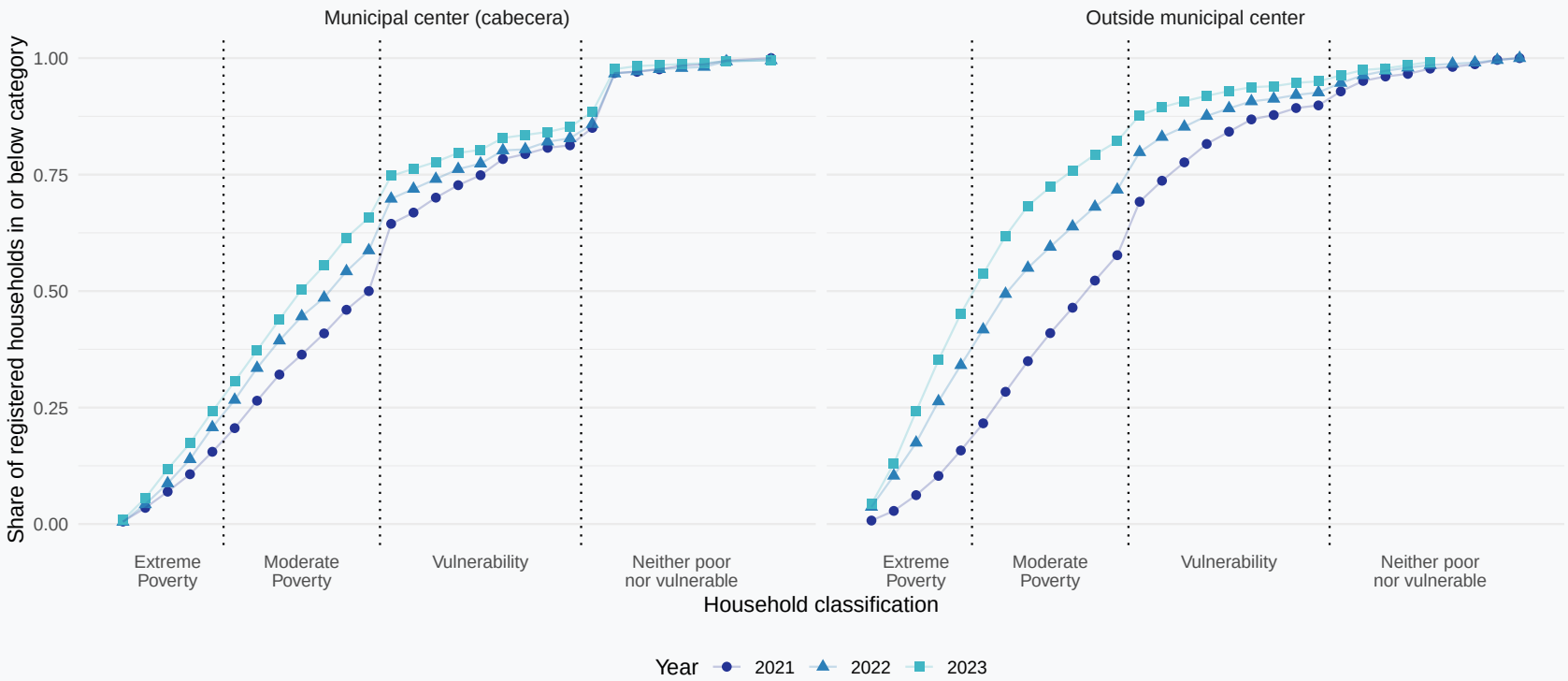
$$\mid Y_{mtc} - \overline{Y}_{mc}^{Pre} \mid \tag{5}$$

Sample municipal SISBÉN-IV classification



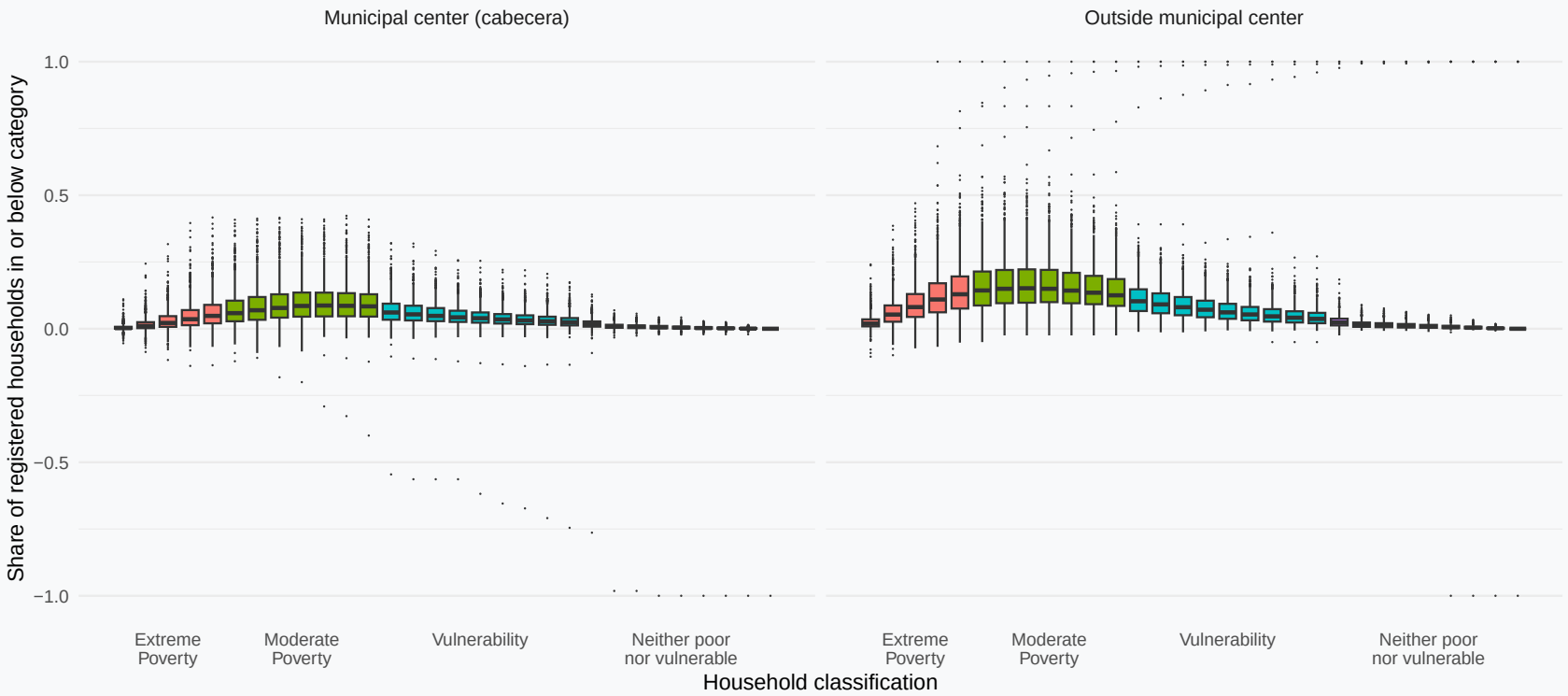
Research design

Cumulative mass functions in one municipality



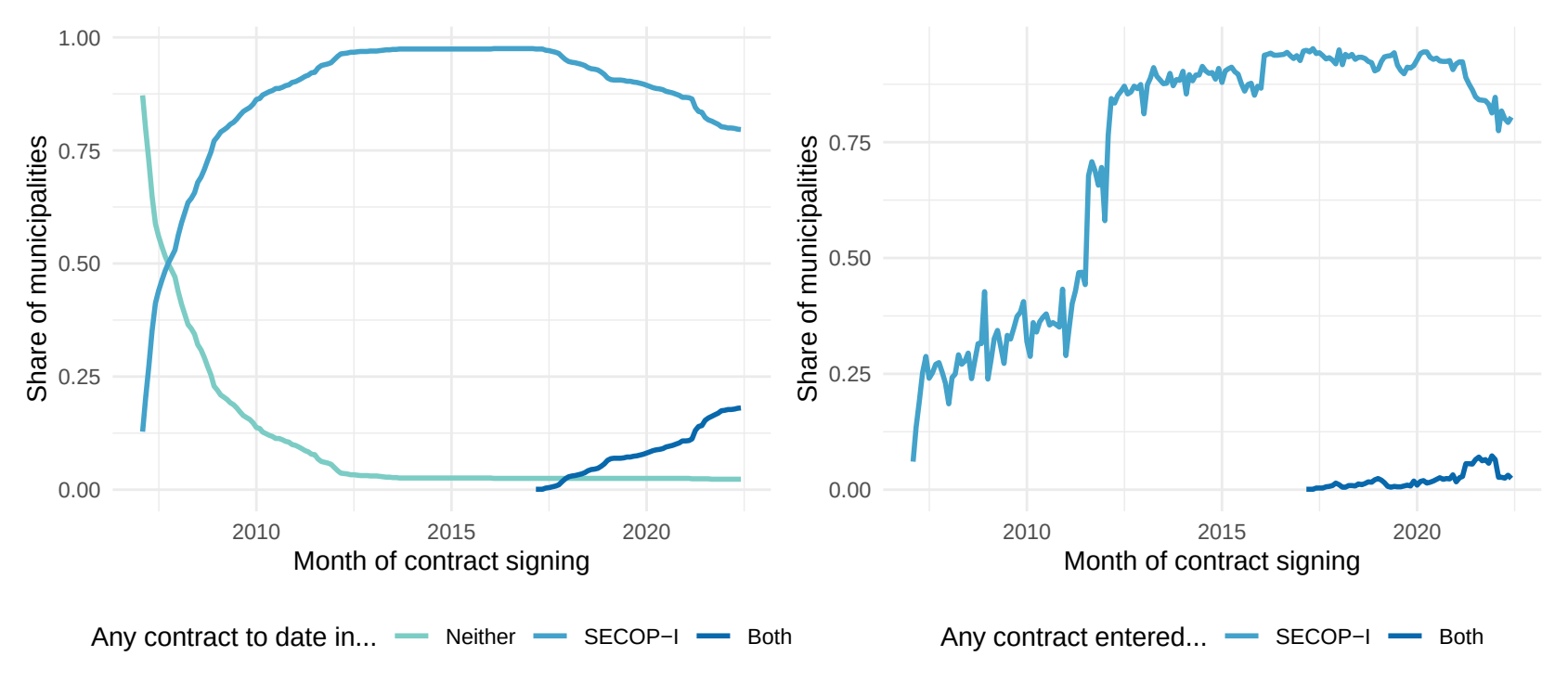
SISBÉN results

Average change in classification, 2021-2022



SISBÉN results

SECOP use



Research design