

The Effects of Increasing Statistical Capacity

Evidence from a Statistical Experiment in Uganda

Eduardo Campillo Betancourt

Northwestern University

19/02/2020

Motivation

- Developing country governments often do not have adequate information about their citizens. (Liaqat, 2020)
- A natural channel through which governments learn is by gathering statistics on their population.
- Improving a country's statistical capacity is a precondition for state-building [Scott (1998); Lee & Zhang (2016)]
- The information gathered can then be used to make policy choices

This paper

Research Question:

- What is the effect of increasing a country's statistical capacity on the allocation of public services?

This paper

Research Question:

- What is the effect of increasing a country's statistical capacity on the allocation of public services?

Empirical Challenges:

- 1) Statistical capacity is endogenous
 - 2) Fundamental challenge in observing a comparison group
- I use a statistical experiment in Uganda to address these concerns

Literature Review

- How state capacity is built and its effects on development [Besley & Persson (2009, 2010); Acemoglu & Robinson (2012); Sánchez de la Sierra (2019)]
- Theory of statistical capacity (legibility) from Political Sciences [Scott (1998); Lee & Zhang (2016)]
- Information interventions involving the government [Liaqat (2020, WP); Hjort et al. (2019)]

Preview of Findings

- The quality and availability of public goods is affected
- No effect on large infrastructure investments or state presence
- Effect is concentrated among relatively simpler improvements and in rural areas

Setting

- Uganda National Panel Survey (a.k.a. Uganda LSMS-ISA), marginal change
- Three **panel** waves since 2009
- ★ Entities and households surveyed are randomly selected ▶ Sampling
- † For waves 4 and 5 (2013 and 2015), a third of the sample was randomly replaced with new households **and** villages.
- Jointly, ★ and † solve the two empirical problems above, where:
 - **Treatment**: Original two thirds of the sample
 - **Control**: Newly added third of the sample

Data

- Observations can be geographically identified at the village-level.
- The outcomes of interest span the following topics:
 - State presence: communication, police, army
 - Provision of health services
 - Investments in infrastructure (roads)
- Data covers 3,119 households across 288 villages
- 67 % of households and 66% of villages are treated

Identification Strategy

The econometric model will be of the form:

$$Y_i = \alpha + \beta \text{treat}_i + \delta_s + \varepsilon_i \quad (1)$$

- Y_i : Outcomes where $i \in \{(h, v), v\}$ for household (h) in village (v)
- δ_s : Sampling strata fixed effects (region \times rural/urban)
- **Treatment**: having been surveyed in the UNPS for the past three waves
- **Control**: just included in the survey, too recent to have had impact

Identification:

$$\mathbb{E}[\varepsilon | \text{treat}, \delta] = 0$$

→ Implied by the random assignment of treatment.

► Balance checks

Results

Health Outcomes

Village-level

Y_v : How has the situation of drinking water in the village changed?

VARIABLES	(1) Weakly better H2O {0, 1}	(2) Weakly better health [0, 1]	(3) Health problems gov't fault [0, 1]
Village surveyed before	0.0849* (0.0695)	-0.0198 (0.605)	-0.0270** (0.0462)
Constant	0.820*** (0)	0.787*** (0)	0.979*** (0)
Observations	283	245	245
Rand. p-value	0.0450	0.592	0.0910

Robust pval in parentheses

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

▶ Household-level

▶ Details on health

▶ Details on government

Health Outcomes

Village-level

Y_v : How have the main challenges of health provision changed?

VARIABLES	(1) Weakly better H2O {0, 1}	(2) Weakly better health [0, 1]	(3) Health problems gov't fault [0, 1]
Village surveyed before	0.0849* (0.0695)	-0.0198 (0.605)	-0.0270** (0.0462)
Constant	0.820*** (0)	0.787*** (0)	0.979*** (0)
Observations	283	245	245
Rand. p-value	0.0450	0.592	0.0910

Robust pval in parentheses

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

▶ Household-level

▶ Details on health

▶ Details on government

Health Outcomes

Village-level

Y_v : Are these challenges to health services under the government's control?

VARIABLES	(1) Weakly better H2O {0, 1}	(2) Weakly better health [0, 1]	(3) Health problems gov't fault [0, 1]
Village surveyed before	0.0849* (0.0695)	-0.0198 (0.605)	-0.0270** (0.0462)
Constant	0.820*** (0)	0.787*** (0)	0.979*** (0)
Observations	283	245	245
Rand. p-value	0.0450	0.592	0.0910

Robust pval in parentheses

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

▶ Household-level

▶ Details on health

▶ Details on government

Infrastructure Outcomes

Village-level

VARIABLES	(1) Rating of road maintenance (1–10)	(2) New roads built ({0, 1})	(3) Access to any road ({0, 1})
Treatment	0.539* (0.0944)	0.0646 (0.328)	-0.0594 (0.239)
Constant	5.777*** (0)	0.690*** (0)	0.796*** (0)
Observations	216	216	288
Rand. p-value	0.0900	0.320	0.230

Robust pval in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Heterogeneity (urban vs. rural)

Village-level

VARIABLES	(1) Weakly better H2O	(2) Health problems gov't fault	(3) Maintenance rating	(4) New roads built	(5)	(6)	(7)	(8)
Treatment	0.0849* (0.0692)	0.0975* (0.0779)	-0.0270** (0.0417)	-0.0356** (0.0274)	0.539* (0.0927)	0.606* (0.0995)	0.0646 (0.332)	0.0388 (0.602)
Urban × Treat		-0.0474 (0.646)		0.0399* (0.0956)		-0.383 (0.584)		0.146 (0.375)
Constant	0.820*** (0)	0.822*** (0)	0.979*** (0)	0.978*** (0)	5.777*** (0)	5.798*** (0)	0.690*** (0)	0.682*** (0)
Observations	283	283	245	245	216	216	216	216

Robust pval in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

► State Presence

► NGO Presence

► Household-level

Conclusion

- Able to overcome the empirical challenges of studying the effects of statistical capacity on local development
- Government becomes more responsive after learning about its citizens
- Results consistent with a government constrained on both funds and information
- The effect is concentrated on rural areas: harder to reach and monitor
- Next steps:
 - Is the allocation of public goods more efficient?
 - Dynamics of the effects with subsequent rounds of the survey
 - Effect at the country level using rollout of Demographic and Health Surveys

Thank you!

Sampling Details

Sampling procedure:

1. Strata defined at the region by urban/rural level (eight strata)
2. Within each stratum, enumerating areas (EAs, villages) were randomly sampled to obtain a nationally- and strata-representative sample
3. For the 2013/14 wave, two thirds of the EAs (and the respective households) were chosen to remain in the panel.
4. The third of the new wave's sample was chosen from the frame developed for the 2014 Population Census

Detailed description

Weakly better health services

Main problems:

- Shortage of personnel
- Shortage/inadequacy of drugs
- Inadequate facilities
- Inadequate funding
- Low pay to staff

◀ Health outcomes

Detailed description

Health problems pertaining the government

Main problems not pertaining to the government:

- Low staff morale
- Negative attitudes of service users
- Decrease in patients
- “Dependency syndrome among patients”
- Bad weather
- Communication problems

Detailed description

Weakly better condition of roads

Main problems:

- Inadequate funding
- Delayed remittance of funds
- Inadequate facilities
- Lack of people's interest
- Inadequate staff
- Wide road network

Detailed description

Road problems pertaining the government

Main problems not pertaining to the government:

- Bad weather
- Bad terrain
- Lack of people's interest
- Hard to reach areas

◀ Infrastructure outcomes

NGO Activity

VARIABLES	(1) # NGOs Total	(2) # NGOs Natl.	(3) # NGOs Intl.
Village surveyed before	0.0842 (0.555)	0.120 (0.297)	-0.0267 (0.406)
Constant	0.456*** (0.000214)	0.390*** (5.64e-05)	0.0498* (0.0777)
Observations	189	189	189
Rand. p-value	0.533	0.293	0.322

Robust pval in parentheses

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

◀ Heterogeneity

“Balance” Checks

- Households remaining in Wave 4 are not positively selected [◀ See table](#)
- Treatment is uncorrelated with “less changeable” variables at the household-level [◀ See table](#)
- Treatment households (and villages) are no more likely to respond to questions [◀ See table](#)
- Effect does not seem to come from the private sector [◀ See table](#)

[◀ Identification Strategy](#)

"Balance" Checks

Selection into treatment

VARIABLES	(1) Weakly better H2O	(2) Health problems gov't fault	(3) Rating of road maint.	(4) # Times reached out to gov't	(5) # Times gov't reached out
Treatment	0.0176 (0.725)	0.0176 (0.506)	0.609 (0.150)	0.232 (0.319)	0.136 (0.557)
Constant	0.898*** (0)	0.796*** (0)	5.563*** (0)	1.407*** (0)	1.377*** (0)
Observations	175	244	130	184	184
R-squared	0.108	0.076	0.188	0.320	0.315
Rand. p-value	0.679	0.516	0.195	0.354	0.580

Robust pval in parentheses

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

◀ Balance

"Balance" Checks

Balance at the household-level

VARIABLES	(1) Primary educ. (father)	(2) Primary educ. (mother)	(3) Baganda ethnicity
HH surveyed last wave	0.0156 (0.395)	-0.0121 (0.507)	-0.00809 (0.423)
Constant	0.376*** (0)	0.604*** (0)	0.187*** (0)
Observations	3,098	3,090	3,118
Rand. p-value	0.396	0.477	0.453

Robust pval in parentheses

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

◀ Balance

“Balance” Checks

Quality of data

VARIABLES	(1) Proportion of NAs in row	(2) Proportion of NAs in row	(3) No rows with NAs	(4) No rows with NAs
Treatment	0.0304 (0.177)	0.00651** (0.0105)	-0.0306 (0.279)	-0.0280 (0.118)
Constant	0.149*** (0)	0.0582*** (0)	0.0724*** (0.00455)	0.503*** (0)
Observations	288	3,119	288	3,119
Rand. p-value	0.183	0.0180	0.235	0.133

Robust pval in parentheses

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

◀ Balance

"Balance" Checks

Private sector

VARIABLES	(1) Priv. health clinic	(2) Priv. hospital	(3) Priv. primary	(4) Priv. secondary	(5) Bank
Village surveyed before	0.0805 (0.168)	-0.0538 (0.367)	-0.0317 (0.489)	-0.0912* (0.0653)	0.0808* (0.0694)
Constant	0.561*** (0)	0.492*** (0)	0.800*** (0)	0.845*** (0)	0.806*** (0)
Observations	285	285	285	284	285
R-squared	0.159	0.149	0.233	0.127	0.210
Rand. p-value	0.142	0.355	0.492	0.0760	0.0540

Robust pval in parentheses

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

◀ Balance

Health Outcomes

Household-level

VARIABLES	(1) Slept under net ({0, 1})	(2) Net treated ({0, 1})
HH surveyed last wave	-0.00226 (0.916)	0.0500*** (0.00254)
Constant	0.589*** (0)	0.863*** (0)
Observations	3,117	2,321
Rand. p-value	0.893	0

Robust p-value in parentheses
Standard errors clustered at the parish level

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

► Village-level

State Presence

Village-level

VARIABLES	(1) Army detach/ barracks	(2) Agricultural Ext. Services	(3) Police station	(4) # Times reached out to gov't	(5) # Times gov't reached out	(6) Prop. of problems solved
Treatment	-0.0103 (0.839)	0.0871 (0.119)	0.00926 (0.571)	-0.414* (0.0652)	-0.337 (0.111)	0.0173 (0.739)
Constant	0.221*** (1.16e-07)	0.658*** (0)	0.980*** (0)	2.049*** (0)	2.057*** (0)	0.478*** (0)
Observations	285	285	285	288	288	224
Rand. p-value	0.826	0.126	0.572	0.0460	0.104	0.760

Robust pval in parentheses

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

◀ Heterogeneity

Heterogeneity (urban vs. rural)

Household-level

VARIABLES	(1) Slept under net	(2) Slept under net	(3) Net treated	(4) Net treated
Treatment	-0.00679 (0.621)	-0.00718 (0.679)	0.0500*** (0.00254)	0.0437** (0.0135)
Urban \times Treat		0.00136 (0.959)		0.0225 (0.570)
Constant	0.794*** (0)	0.794*** (0)	0.863*** (0)	0.863*** (0)
Observations	2,321	2,321	2,321	2,321

Robust p-value in parentheses
Standard errors clustered at the parish level
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

► Village-level