

UNBLINDED RESULTS

Third-Party Performance Pay to Improve Local Government Accountability: A Field Experiment in Burkina Faso

Malte Lierl (German Institute for Global and Area Studies / GIGA)*

Marcus Holmlund (World Bank, Development Impact Evaluation / DIME)

February 6, 2023

Abstract

Can local government accountability be improved by giving community-based organizations a financial stake in their local government's performance? In a field experiment in Burkina Faso, we test a "third-party performance pay" scheme for community-based organizations (CBOs). Selected CBOs are promised a variable cash grant that is proportional to changes in their local government's performance scores over a two-year period. We test if third-party performance pay (1) motivates CBOs to actively lobby for better municipal performance, (2) increases accountability and problem-awareness of municipal decision makers and (3) ultimately leads to improvements in municipal government performance. We also investigate if the incentive scheme had any unintended consequences for the internal functioning of the beneficiary CBOs.

***Note to readers:** This document is the unblinded replication of a results-blind analysis report that was previously uploaded to the AEA RCT registry. The blind analysis was conducted using datasets where all variables indicating treatment assignment or treatment status had been removed, masked, or replaced with simulated (randomly permuted) treatment identifiers. The original datasets were encrypted and safeguarded by research team members who are not authors of this study. During the blind analysis, we developed and refined our analytical framework, data cleaning and estimation strategies with knowledge of the data, but without knowledge of the actual results. Furthermore, we incorporated results-blind expert feedback from seminar presentations and recorded participants' expectations about the eventual, unblinded results via a prediction survey.

*Corresponding author. Contact: malte.lierl@giga-hamburg.de

Contents

1	Introduction	4
2	Experimental Intervention	6
2.1	Selection of partner CBOs	6
2.2	Reward scheme and performance metrics	7
2.3	Information and training	9
2.4	Design considerations	9
2.5	Implementation	9
3	Theoretical Expectations and Research Hypotheses	10
3.1	Hypothesis 1: A financial stake in municipal government performance increases CBOs' involvement in municipal governance.	12
3.2	Hypothesis 2: 3PPP does not erode the internal functioning or collective action capacity of CBOs. . . .	12
3.3	Hypothesis 3: 3PPP causes municipal decision makers to perceive greater accountability pressure	13
3.4	Hypothesis 4: 3PPP increases municipal decision makers' awareness of performance metrics	13
3.5	Hypothesis 5: 3PPP increases municipal performance	13
4	Blind Analysis Methodology	14
4.1	Blinded datasets	14
4.2	Blind analysis process	15
5	Treatment Assignment, Compliance and Attrition	17
5.1	Treatment assignment and compliance at the municipality level	18
5.2	Treatment assignment at the CBO level	19
5.3	Attrition and non-response at the CBO level	20
5.4	Attrition and non-response at the level of municipal decision makers	20
5.5	Balance tests	20
6	Estimation Strategy and Results-Blind Analysis	21
6.1	Hypothesis 1: A financial stake in municipal government performance increases CBOs' involvement in municipal governance.	21
6.1.1	Outcomes of interest and data sources for Hypothesis 1	21
6.1.2	Estimation strategy for Hypothesis 1	25
6.1.3	Unblinded results for Hypothesis 1	25
6.1.4	Alternative outcome indicators and estimation strategies considered during blind analysis	27
6.2	Hypothesis 2: 3PPP does not erode the internal functioning or collective action capacity of CBOs. . . .	29
6.2.1	Outcomes of interest and data sources for Hypothesis 2	29
6.2.2	Estimation strategy for Hypothesis 2	33
6.2.3	Unblinded results for Hypothesis 2	33
6.2.4	Alternative outcome indicators and estimation strategies considered during blind analysis	34
6.3	Hypothesis 3: 3PPP causes municipal decision makers to perceive greater accountability pressure. . . .	35
6.3.1	Outcomes of interest and data sources for Hypothesis 3	35
6.3.2	Estimation strategy for Hypothesis 3	35
6.3.3	Unblinded results for Hypothesis 3	37
6.3.4	Alternative outcome indicators and estimation strategies considered during blind analysis	37
6.4	Hypothesis 4: 3PPP increases municipal decision makers' awareness of performance metrics.	38

6.4.1	Data and methods for Hypothesis 4	38
6.4.2	Estimation strategy for Hypothesis 4	39
6.4.3	Unblinded results for Hypothesis 4	39
6.4.4	Alternative outcome indicators and estimation strategies considered during blind analysis	40
6.5	Hypothesis 5: 3PPP increases municipal performance.	41
6.5.1	Outcomes of interest for Hypothesis 5	41
6.5.2	Estimation strategy for Hypothesis 5	41
6.5.3	Unblinded results for Hypothesis 5	42
7	Expectations Survey	44
8	Conclusions	45
9	References	46

1 Introduction

In many parts of the world, local-level governments struggle to provide basic public services to citizens. Often, local governments' capacities are limited by material constraints, such as inadequate funding, insufficient skills and human resources, poor information, or a lack of basic equipment and infrastructure. However, merely alleviating these material constraints may be insufficient to increase local government performance, if local governments are not also accountable for their performance. For that reason, governments, NGOs and development organizations are investing in interventions to improve the accountability of local governments and service providers.

Interventions to improve local government accountability can broadly be classified into top-down and bottom-up approaches. Top-down approaches tend to involve external monitoring and performance-linked incentives, for example in the form of results-based financing or pay-for-performance schemes. Bottom-up approaches rely on public pressure from citizens. Interventions in the latter category include strategies to increase citizen participation (for example through participatory budgeting or community-driven development), information provision to citizens (for example through voter information campaigns), or combinations of both (for example community monitoring interventions). In each category, there have been some successes and some failures and certain limitations have become apparent.

Top-down monitoring and performance incentives, such as results-based financing and pay-for-performance, have been widely applied at the level of service providers, especially in the health and education sectors. Some studies suggest that pay-for-performance schemes can improve service delivery, but the overall evidence is mixed (Asadi-Aliabadi et al., 2022; Diaconu et al., 2021; Singh et al., 2021). Criticism of top-down performance incentives comes both from public policy perspectives (Paul et al., 2018; Paul et al., 2021), as well as from behavioral economists, who emphasize that performance rewards can adversely affect the intrinsic motivation of service providers (Deci et al., 1999; Fang and Gerhart, 2012; Frey, 2017), their image motivation (Ariely et al., 2009; Belle and Cantarelli, 2015), and peer pressure to perform (Fuster et al., 2010).

With regard to bottom-up accountability interventions, a growing body of evidence suggests that their impact strongly depends on program design and local context. Information provision alone may fail to increase the accountability of local governments, because voters do not necessarily sanction poor performance at the polls (Dunning et al. 2019). Externally induced citizen participation, on the other hand, tends to be costly and vulnerable to known sustainability problems (Mansuri and Rao, 2012), such as elite capture (Platteau and Gaspard, 2003; Ensminger, 2012) and excessive demands on citizens (Khwaja, 2004). If external facilitation and funding flows are removed, "induced" participation typically does not continue on its own (Mansuri, 2013). Some success has been achieved by community-monitoring programs that combine information provision with external facilitation of citizen participation (e.g. Björkman and Svensson, 2009; Barr et al., 2012; Nyquist et al. 2017; Andrabi et al., 2017; Fiala and Premand, 2018; Christensen et al., 2021). However, these successful instances contrast with other cases where community monitoring had little or no positive impacts on public service provision (Olken 2007; Banerjee et al. 2010; Raffler et al. 2022).

In this paper, we introduce an alternative approach to promoting local government accountability, which we call "third-party performance pay" (3PPP). 3PPP combines elements of top-down incentives and bottom-up citizen pressure. It relies on existing, well-functioning local collective action groups (community-based organizations or CBOs), and incentivizes them to hold local governments accountable for their performance. This is accomplished by giving the CBOs a direct financial stake in the performance of their local government. An unconditional cash grant is promised to them that is proportional to improvements in their municipal government's performance indicators over a two-year period. Until the end of the evaluation period, the CBOs can potentially increase their payoff by taking action to increase their local government's performance. However, since the cash grants are not conditional on any specific actions the CBOs undertake – even if the CBOs do nothing at all, they remain entitled to the cash grant – it incentivizes only those actions the CBOs believe to be worth their time and effort and that are sufficiently likely to have an actual impact on municipal government performance.

We test third-party performance pay (3PPP) at the municipal level in a nationwide randomized controlled trial in Burkina Faso. The experiment was implemented, in collaboration with the government of Burkina Faso, in 149 treatment and 150 control municipalities. It involved 149 treatment CBOs (in treatment municipalities only) and 449 control CBOs (in both control and treatment municipalities). We examine if 3PPP motivates CBOs to actively lobby for better municipal performance, if it increases accountability pressures on municipal decision makers, and if it ultimately leads to improvements in municipal government performance. We also investigate if the incentive scheme had any unintended side effects on the functioning of beneficiary CBOs.

Our research is of theoretical as well as practical relevance, as 3PPP differs in key aspects from existing and previously studied approaches to promoting the accountability of local governments and public service providers.

First, 3PPP has the potential to avoid the adverse consequences of direct financial incentives to public servants, especially with regard to their intrinsic motivation and with regard to concerns about the inequitable treatment of service providers who face external constraints and hardships that are beyond their control. 3PPP circumvents these issues by avoiding financial incentives to service providers altogether. For CBOs, in turn, the rewards are not tied to any specific actions they undertake. Furthermore, they are too limited in size to cover the costs of actually engaging in service provision on behalf of the municipality. Therefore, 3PPP primarily incentivizes catalytic activities, such as collaborative problem-solving with municipal decision makers, or mobilizing citizens to exert public pressure. These activities still require a great deal of intrinsic motivation and voluntary collective action from everyone involved. Therefore, we expect that 3PPP might be less damaging to the intrinsic motivation of CBOs than other types of results-based incentives.

Second, 3PPP eliminates two controversial aspects of community monitoring programs: (1) The costly creation of new local institutions or collective action structures through external facilitation and injection of funds, and (2) the power imbalance between local actors and external funders, which may lead to the underutilization of local knowledge or cause local actors to waste their effort on activities that may turn out to be suboptimal in the local context. Unlike community monitoring programs, which require large up-front investments and continued support to create effective local accountability structures (Banerjee et al., 2010), we leverage the already existing collective action capacity and social influence of local CBOs that are invited to apply for the incentive scheme and competitively selected. Furthermore, 3PPP does not prescribe any specific activities to the selected CBOs. Partner CBOs decide on their own if, how and to what extent they want to try to influence the performance of their municipal government. Since cash grants to partner CBOs are not conditional on any specific actions, and partner CBOs are eligible for a cash grant even if they do nothing at all, they are incentivized to only undertake activities they find worthwhile, based on their local knowledge. Consequently, we expect that 3PPP might result in potentially less wasted effort and potentially superior value for money compared to traditional community monitoring approaches.

Third, 3PPP differs from the contracting of civil society organizations to deliver specific services at the local level, as it is frequently practiced by development programs who rely on civil society organizations as implementation partners. Rather than being paid to perform specific actions or instructed to follow certain “best practices”, 3PPP encourages the partner CBOs to develop their own strategies of influence, using their local knowledge and the resources available to them. By collecting data on the strategies that CBOs pursue out of their own motivation, our project also sheds light on mechanisms of informal accountability and civil society influence in local governance and on the underlying social and political norms.

Finally, 3PPP has the potential to mitigate incentive problems that prevent high-functioning CBOs from contributing more actively to local-level governance. Often, community-based organizations (CBOs) are among the highest-functioning entities at the local level. Some CBOs excel at mobilizing communities and facilitating collective action, for example hometown associations and womens’ groups. Others have substantial financial management capacity, such as producer cooperatives, savings associations and self-help groups. Yet others fulfill important roles in the provision of public goods, such as parent-teacher associations and user groups of common-pool resources. However, despite their capabilities,

which sometimes compare favorably to those of local governments, most CBOs participate in local governance only in limited ways, because they primarily exist to serve the interests of their member base, rather than of the community as a whole. If high-functioning local CBOs had a more comprehensive interest in local governance – and were rewarded for the efforts rendered on behalf of the community as a whole – they could, by virtue of their social capital and local knowledge, be an important source of advice and accountability for local governments.

The paper proceeds as follows. We first describe the experimental intervention in detail. Next, we describe our theoretical expectations and research hypotheses, as well as the results-blind analysis methodology we use to shield our study from experimenter bias. This is followed by a description of the experimental design and implementation. We then provide a detailed description of the outcome data, estimation strategies and **unblinded** results, along with notes about alternative approaches considered and judgments taken during the results-blind analysis. Finally, we present survey results regarding the effect sizes expected by researchers to whom we presented **the** results-blind report. Upon unblinding of the study, we will fill in the actual results, discussion and conclusions.

2 Experimental Intervention

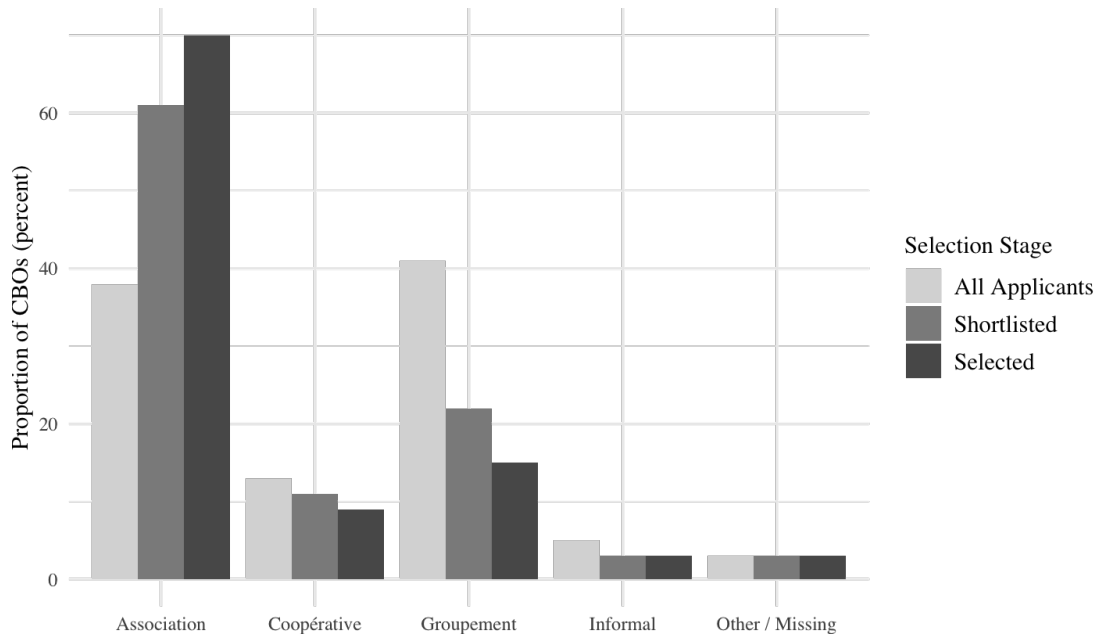
In collaboration with Burkina Faso’s Territorial Collectivities Support Program (*Programme d’appui aux collectivités territoriales* or PACT), we developed and tested a third-party performance pay (3PPP) scheme for municipalities between 2013 and 2021. In 2016, a small-scale proof-of-concept trial was carried out in six randomly selected municipalities and a qualitative study was conducted to assess the feasibility of the approach and to gain insights for its refinement. Upon encouraging findings from the proof-of-concept study, 3PPP was then implemented as a nationwide randomized controlled trial, under the program title “PACT Partnerships with Community-Based Organizations”. The trial covered 149 treatment and 150 control municipalities between 2019 and 2021.

Under the 3PPP scheme, we selected two of the best-functioning and most influential local community-based organizations (CBOs) in every municipality, following a public call for applications. In treatment municipalities, one of the two selected CBOs was randomly chosen as partner CBO and promised a variable cash grant. The amount of this cash grant was to be proportional to changes in their municipal government’s performance scores over a two-year period, as measured independently through Burkina Faso’s annual municipal performance monitoring survey (SUPERMUN). The usage of the cash grants by CBOs was unrestricted and the amount depended solely on the municipal government’s performance scores. Thus, partner CBOs were given a direct financial stake in the future performance of their municipal government. The CBOs were not contracted for any specific work – they would get paid even if they did nothing. However, they could potentially increase their payoff by nudging municipal decision makers to address shortfalls in municipal service delivery and institutional capacity, if they believed that doing so would be worth their time and effective at increasing their municipality’s performance.

2.1 Selection of partner CBOs

To select treatment and control CBOs, an identical public call for applications was organized in all treatment and control municipalities. This call for applications was disseminated through a combination of public announcements, local information sessions and advertisements on local radio. Local collective action groups of different legal forms, including associations, cooperatives, interest groups and informal organizations were invited to apply, under the condition that they had to be local, membership-based grassroots organizations. Large national or international NGOs, private-sector enterprises, political parties or religious communities of worship were ineligible. Figure 1 breaks down the applicant CBOs by type of organization and selection outcome. In total, 4,749 expressions of interest from CBOs were received from across 331 municipalities. 402 applications were from 32 municipalities that later had to be dropped from the experiment due to security concerns, leaving 4,347 applications from the 299 municipalities that were ultimately included in the study.

Figure 1: Types of CBOs by selection stage



The application procedure and the review of applications was blind to the treatment status of the municipality, so that the exact same selection procedure was applied in treatment and control municipalities. A municipality’s assigned treatment status was disclosed to our implementing partners only after the CBO selection process in that municipality had been completed. A first screening ensured that only local community-based organizations were considered, not private-sector firms, political parties, religious communities or large national or international NGOs. The first eligibility screening was followed by a scoring along a set of quantitative selection criteria (Table 1): the number of active members in the CBO, the annual budget, the number of years since the founding of the CBO, the number of villages of the target municipality in which the CBO had members, the number of ethnic groups represented within the CBO, the highest number of people the CBO has been able to mobilize for a public event, and the extent to which the CBO had previously interacted with the mayor or the secretary-general of the municipality. On average, 4.7 CBOs per municipality were shortlisted on the basis of these quantitative selection criteria. The shortlisted CBOs were additionally scored on a qualitative application section, which consisted of open-text statements on their motivation and experience. The shortlisted CBOs were ranked based on their combined scores on the quantitative and qualitative application sections. The two top-ranked CBOs in every municipality were retained for random assignment into treatment and control groups.

2.2 Reward scheme and performance metrics

Partner CBOs were promised a cash grant that was proportional to changes in their municipal government’s performance scores over a two year period. Municipal government performance scores are based on Burkina Faso’s annual municipal performance survey, SUPERMUN (*suivi de la performance municipale*). SUPERMUN data is collected annually, combining multiple administrative data sources and facility surveys. SUPERMUN has been tracking sixteen indicators of service delivery and institutional capacity in six regions since 2013 and nationwide since 2017. The SUPERMUN indicators were developed through extensive consultations with mayors, municipal administrators, government agencies and local-level service providers. They focus on the most common and most consequential bottlenecks in municipal government performance in the areas of primary education, health care, water and sanitation, administrative services, municipal staffing, tax collection, procurement, and governance. For example, municipal governments are in charge of delivering school supplies, such as textbooks, to primary schools at the beginning of the school year. In some

	All Applicants	Mean/Proportion Shortlisted	Selected
Active members	99.91	159.96	229.82
Years of existence	8.53	10.44	10.99
Has formal budget	0.75	0.78	0.78
Average annual budget (in million FCFA)	2.2	4.38	6.71
Number of villages in which the CBO is active	7.01	13.46	15.71
Number of ethnic groups represented ⁱ	2.61	3.51	3.72
Mobilization capacity ⁱⁱ	601.86	972	1589.39
Prior involvement in municipal affairs ⁱⁱⁱ	0.65	0.92	0.87
<hr/>			
Number of CBOs	4749	1537	660

ⁱ⁾ Among the CBO's membership base in the commune where the CBO is applying.

ⁱⁱ⁾ The greatest number of people ever mobilized by the CBO for an event in their commune, according to the CBO's application form.

ⁱⁱⁱ⁾ The CBO has met with the mayor or municipal secretary-general in 2018 to discuss local issues.

Table 1: Quantitative selection criteria

municipalities, these deliveries are delayed by weeks or months, limiting teachers' ability to teach, with drastic consequences for learning outcomes. Similarly, primary health centers depend on regular delivery of gas bottles from municipalities for their refrigeration systems. Gas stockouts are common and interrupt the refrigeration chain, which causes vaccines and essential medications, such as insulin, to spoil. In addition to their relevance, SUPERMUN indicators were selected to be traceable over time and secondarily to be comparable across municipalities. All SUPERMUN indicators have target values that correspond to national standards for service delivery or institutional capacity which most municipalities fail to meet. Thus, SUPERMUN performance indicators are particularly adept at highlighting shortfalls in municipal government performance and institutional capacity that ought to be remedied, because they have outsized negative consequences for the local communities.

Based on the performance indicators, SUPERMUN assigns an overall performance score to municipalities. The scoring algorithm was designed in collaboration with our government partners and assigns points to indicator values based on their marginal importance to the central government's development objectives and the expected marginal effort needed for improvements from different levels. The scoring function is non-linear in most indicator values but monotonic with respect to all indicators. Better indicator values always correspond to higher performance scores up to the point at which the municipality has reached its target. The overall municipal performance scores can be interpreted as a metric of how close the municipality is to national performance standards on a set of indicators that reflects the central government's priorities at the time of the study.

The cash grants promised to CBOs were not conditional on any specific actions by the CBOs. They depended solely on changes in their municipal government's performance score from a baseline period (2018) to an evaluation period (2020). The payoffs were calculated as follows:

$$\text{Payoff}_{2021} = 10000 \text{ FCFA} \cdot \max(0, 50 + 6 \cdot (\text{Performance}_{2020} - \text{Performance}_{2018}))$$

The CBOs received their cash grants in mid-2021, after the municipal performance data for 2020 had been collected and processed. No restrictions were imposed on how the CBOs could use their cash grants, and there were no reporting requirements. To receive the transfer, the CBOs merely had to provide their bank account information to the sponsor and were contacted to ensure the money had been properly received. CBOs that had no bank account were asked in advance to designate a person, e.g. a treasurer, who would receive the transfer on behalf of the organization, for

example via mobile money.

2.3 Information and training

Upon selection of the partner CBOs, a meeting was organized for each treatment CBO to explain the rules of the program, the responsibilities of municipal administrations, how the municipal performance metrics are calculated, and the modalities of the cash grant. The timing of this meeting was chosen to coincide with the dissemination of SUPERMUN performance scorecards to municipal decision makers, which took place both in the treatment and in the control municipalities. In the treatment municipalities, the treatment CBOs were also invited to attend this dissemination event, introduced to municipal decision-makers at that occasion and provided with copies of their municipality's performance scorecard. Following the scorecard dissemination event, the partner CBOs were encouraged to brainstorm if and how they might be able to induce municipal decision makers to address performance shortfalls.

2.4 Design considerations

A core design principle for the 3PPP approach was to avoid unintentional adverse impacts on the participating CBOs or on the intrinsic motivation of their members. For that reason, the cash grant scheme imposed only minimal requirements on the participating CBOs. Instead of being told what to do, CBOs were merely given a financial stake in the performance of their municipal government and encouraged to find their own ways of influencing their municipal government's performance. The program itself was as hands-off as possible, while providing partner CBOs with the information necessary to make an informed decision about their participation.

Since the cash grants were not conditional on any specific actions by the CBOs, they were likely to incentivize only those actions that the CBO believed would have enough of an impact on their payoff to make it worth the effort. Unlike contracting CBOs for specific activities, which might induce dependency or distract CBOs from their original mission, 3PPP leaves the decision to what extent the CBO wants to get involved in municipal governance with its members and gives CBOs the flexibility to freely adjust their involvement at any time, as circumstances change. The idea is that this hands-off approach would minimize wasted efforts on activities or forms of participation that CBO members believe to be futile.

The magnitude of the cash grants and the payoff formula were chosen with the intention of (1) providing a meaningful incentive to CBOs, (2) keeping the payoff small enough that it would not offset the costs of doing work that is the municipality's responsibility, (3) limiting potential financial liability, and (4) preserving scalability. A potential weakness of the payoff formula is that it does not adjust for overall trends in municipal performance. If municipal government performance decreases overall, as it was the case during the study period due to an escalation of armed conflict and economic crisis, the incentive to CBOs is weakened in those municipalities, in which the decrease in performance scores is so severe that corresponding reduction in the CBO's payoff exceeds the fixed based amount of 500,000 FCFA, so that the CBO's payoff is zero and efforts to mitigate the declining municipal government performance would no longer result in increased payoffs to the CBO. Conversely, if municipal performance increases for external reasons, there are hard-to-calculate financial risks for the sponsor. An alternative approach would have been to adjust the payoff for overall trends observed in the control group. In that case, 3PPP would, in expectation, be paying only for the causal effect of the intervention on municipal government performance. However, this proposal was rejected by our government partners over concerns about scalability, because it would no longer be feasible if the intervention were scaled up to cover the control group.

2.5 Implementation

The randomized controlled trial was implemented in collaboration with the Ministry of Territorial Administration, Decentralization and Social Cohesion (MATDCS) of Burkina Faso, specifically its Local Governance Support Program

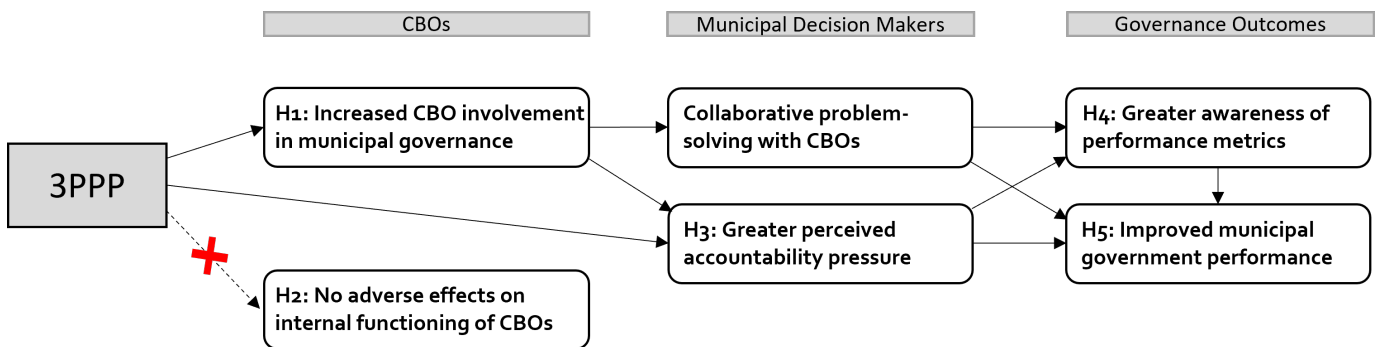
(*Programme d'appui aux collectivités territoriales / PACT*), which was supported by the World Bank. The experiment was intended to contribute to the government's policy objectives of increasing participation of community-based organizations (CBOs) in municipal governance and improving the accountability of municipal decision makers. The idea for the CBO incentive scheme was developed by the authors and pitched, along with other ideas, to various government agencies in 2013. From an initial brainstorming workshop on policy experimentation in decentralization and local governance, a collaborative research and policy learning agenda emerged that involved several randomized-controlled trials, including this experiment. This collaborative research program is titled REGLAB (*recherche expérimentale sur la gouvernance locale au Burkina Faso*).

A national NGO, *Association Mains Unies Pour le Sahel* (AMUS) was contracted as implementation partner responsible for (1) managing the call for applications, (2) disseminating SUPERMUN performance scorecards to the municipal governments in both treatment and control municipalities, (3) training partner CBOs in treatment municipalities, (4) and operating a question hotline for partner CBOs. The intervention manual, as well as all materials and instructions were developed collaboratively between the research team and local stakeholders. The *Centre d'études et d'expérimentations économiques and sociales de l'Afrique de Ouest* (CESAO) provided independent quality control and monitored strict adherence to the intervention manual and experimental protocol. For the endline data collection, we partnered with *Innovations for Poverty Action* (IPA) and CESAO.

3 Theoretical Expectations and Research Hypotheses

Our research objective is to examine the effectiveness of 3PPP, by evaluating its impacts at different steps of the causal chain that links the incentives provided by 3PPP to changes in municipal government performance. We investigate five hypotheses regarding the effect of 3PPP on CBOs' involvement in municipal affairs (H1), on the CBOs' internal functioning (H2), on the accountability pressure perceived by municipal decision makers (H3), on municipal decision makers' awareness of performance shortfalls (H4) and ultimately on actual municipal performance outcomes (H5).

Figure 2: Causal pathways through which 3PPP could affect municipal government performance



Theoretically, we expect that 3PPP could have an effect on municipal government performance through two causal pathways, as detailed in Figure 2. The first causal pathway works by incentivizing cooperation between municipalities and CBOs. For example, CBOs might have constructive conversations with municipal decision makers to identify problems and potential solutions, or use their social capital and standing in the community to facilitate collective action in support of the municipality's goals. Collective problem-solving by municipal governments and CBOs should be reflected in increased involvement of the CBOs in municipal governance (H1).

The second causal pathway works through accountability pressures on municipal decision makers. Knowing that the partner CBOs have a stake in the municipal government's performance, as well as the ability to exert pressure the municipal government if necessary, municipal decision makers feel that under-performing on their responsibilities could

trigger some form of accountability, e.g. in the form of informal social pressure, public protest, or other forms of citizen mobilization (H3). Importantly, for 3PPP to have an effect on perceived accountability pressures, it is not necessary that CBOs take any action. The mere anticipation that the CBO might exert pressure could be sufficient to induce changes in municipal decision makers' perceptions of accountability. If that is the case, then it may be off the equilibrium path for CBOs to actually take any action. Similarly, subtle signaling by CBOs of their willingness to take action could have the desired effect, but might not require much actual activity by the CBOs. In either case, we might see Hypothesis 3 supported, even if there is no evidence in support of Hypothesis 1.

A third potential causal pathway, access to information about performance shortfalls, was held constant in the experiment, because municipal decision makers in the control municipalities were given the same information as those in the treatment municipalities via the SUPERMUN performance scorecards and the local dissemination events. Therefore, we are confident that access to information was highly comparable across treatment and control groups. If we see any differences in municipal decision makers' awareness of performance metrics (H4), those effects should be attributable to greater perceived accountability pressure (which could drive decision makers' attention to performance metrics) or efforts by CBOs to draw municipal decision makers' attention to performance shortfalls, rather than differences in information availability.

If the experiment supports neither Hypothesis 1 nor Hypothesis 3, we would not expect an effect on municipal performance outcomes (H5), as there appear to be no plausible ways in which 3PPP could either increase the intrinsic motivation of municipal decision makers to fix performance shortfalls, or alleviate other types of constraints on municipal government performance. Furthermore, since the promised cash grant to CBOs is not sufficient to cover the costs of CBOs providing services that municipalities fail to provide, we also do not expect that 3PPP could influence municipal performance outcomes through that channel.

Our prior expectations regarding these hypotheses are based on the proof-of-concept pilot study, which we conducted in six randomly selected municipalities in Burkina Faso in 2016. Based on focus group interviews with members and leaders of the partner CBOs, we believe that the 3PPP incentive scheme had a positive motivational effect on five of the six pilot CBOs. These five CBOs were located in rural areas, whereas the sixth CBO was located in an urban area. The five rural pilot CBOs reported to have substantially increased their involvement in municipal governance in various ways as a result of the incentive scheme, engaging in activities such as information gathering, meetings and advocacy with municipal authorities, mobilization of public pressure and co-production with municipal authorities. The results of the proof-of-concept phase were encouraging enough to warrant a randomized-controlled trial on a nationwide scale, which was subsequently implemented. Our hypotheses for the nationwide RCT are based on qualitative insights from this proof-of-concept phase.

While we do expect the proof-of-concept results to be predictive of the impacts of the scaled-up intervention, there are three caveats. First, the precarious security situation in many areas of Burkina Faso as well as the shutdown of public life due to the COVID-19 pandemic have hamstrung municipal governments' ability to deliver on even their most basic responsibilities. In some areas, municipal governments have additionally been affected by displacement of public employees and elected local decision-makers. This added stress on municipal governments might decrease the extent to which CBOs believe they can make a difference in municipal governance. Second, ability of CBOs to carry out meetings, assemblies or other public activities is severely curtailed by COVID-19 and security-related restrictions on public life, such as social distancing measures, curfews, etc. Compared to normal times, CBOs in this experiment have much more limited opportunities and capabilities to get involved in municipal governance. Third, the CBOs' financial incentives have likely been weakened by an overall downward trend in municipal performance and capacity in Burkina Faso, because the payoff formula was designed in such a way that no reward could be earned if municipal performance scores deteriorated by more than a particular threshold. This threshold is likely to have been reached in some municipalities. Therefore, we expect the effects of 3PPP in the experiment to be potentially less pronounced than under more normal circumstances.

In what follows, we will discuss the assumptions and implications of each of the five research hypotheses in detail.

3.1 Hypothesis 1: A financial stake in municipal government performance increases CBOs' involvement in municipal governance.

We hypothesize that 3PPP will increase CBOs' involvement in municipal governance (Hypothesis 1). From a theoretical perspective, this requires that three conditions are satisfied:

- CBOs are motivated by financial incentives. While it is sometimes assumed that organizations respond to financial incentives just like individuals, this assumption may not be self-evident in the case of community-based organizations. Some CBOs might need little or no money to pursue their organizational mission and might therefore not be strongly motivated by the prospect of augmenting their budget, whereas others might greatly benefit from an additional source of revenue.
- The experimental intervention creates an actual financial incentive for the CBOs. This might not be the case in municipalities that experienced a steep decline in performance scores over the study period, such that the partner CBO's payoff would remain zero even if they were able to positively influence municipal government performance.
- The CBOs believe that they can influence municipal government performance in ways that would result in a greater payoff for them. CBOs' ability to influence municipal government performance through their own action may vary due to municipality-level conditions (e.g. how easy it is to resolve certain performance issues) and due to differences in CBOs' resources, costs, and collective action capacity.

Since not all CBOs necessarily satisfy these preconditions, we would expect treatment effects to be heterogenous. Their magnitude should vary with (1) the extent to which a cash grant could actually help CBOs attain their objectives, (2) the presence of adverse external shocks to municipal government performance and (3) the capacity of CBOs to actually influence municipal government performance. The latter should vary with the CBOs' baseline characteristics (e.g. the size of its membership base and their operating budget), the date at which the partnership was launched (i.e. the time remaining to influence municipal performance within the payoff-relevant period). Based on experiences from the pilot test, we further consider the possibility that CBOs in urban municipalities might see fewer opportunities to influence municipal government performance, compared to CBOs in rural municipalities, and may therefore have less of an incentive to get involved in municipal affairs.

3.2 Hypothesis 2: 3PPP does not erode the internal functioning or collective action capacity of CBOs.

With regard to the internal functioning and collective action capacity of the CBOs, we do not expect that 3PPP has any adverse consequences (Hypothesis 2). In principle, the CBO incentive scheme could even reinforce CBOs' collective action capacity: From the perspective of individual CBO members, the cash grants are a collective good that can potentially be augmented through collective action. Hypothetically, however, the incentive scheme could also aggravate internal conflicts within CBOs, because involvement in municipal governance is not among the original objectives of most CBOs. Actions taken in pursuit of this goal could involve a reallocation of efforts and resources within the CBOs. If members of a CBO disagree about the merits and opportunity costs of getting involved in municipal governance, the incentive scheme could unintentionally cause internal conflict within some CBOs. If such internal conflicts are not productively resolved, they could cause tension, frustration and exit from the organization on the part of some members. From a policy perspective, it is therefore important that this experiment also screens for such potential adverse consequences.

3.3 Hypothesis 3: 3PPP causes municipal decision makers to perceive greater accountability pressure

At the level of municipal decision makers, we hypothesize that 3PPP causes municipal decision makers to perceive greater accountability pressure (Hypothesis 3). This hypothesis is based on two potential channels of influence. First, through gradual learning, in the sense that decision makers could adjust their perceptions of accountability after they observe that incentivized CBOs actually hold them accountable for their performance. Second, through the anticipation that the incentivized CBOs would exert pressure on municipal decision makers, if the latter do not sufficiently address shortfalls in municipal government performance. If decision makers respond pre-emptively and address performance issues, they obviate the need for CBOs to actually take any action, but their perception of accountability pressure would have increased nevertheless.

The experimental design does not distinguish between gradual learning and pre-emptive anticipation. However, if anticipated accountability pressure dominates, then we might see effects on municipal decision-makers' perceptions of accountability, even if Hypothesis 1 cannot be confirmed, because CBOs do not actually need to get involved.

3.4 Hypothesis 4: 3PPP increases municipal decision makers' awareness of performance metrics

With regard to municipal decision makers' awareness of performance shortfalls, we hypothesize that 3PPP could increase, but not decrease municipal decision makers' awareness (Hypothesis 4). However, we expect that increases in municipal decision makers' awareness would likely be of limited magnitude, because information about performance shortfalls was disseminated equally to municipalities in the treatment and control groups, using the same performance scorecard that was also presented to the partner CBOs and used to calculate their payoff. Therefore, most municipal decision-makers in both the treatment and control group should already be aware of the SUPERMUN performance metrics.

However, it is possible that decision makers in treatment municipalities pay greater attention to performance metrics, because of greater perceived accountability pressures, because CBOs draw decision makers' attention to them. These circumstances could result in a greater awareness of performance metrics among decision makers in treatment municipalities.

3.5 Hypothesis 5: 3PPP increases municipal performance

With regard to municipal performance scores, we test if 3PPP results in any measureable increases in municipal performance (Hypothesis 5). We expect that 3PPP could increase municipal performance scores, if it effectively motivates CBOs to get involved (Hypothesis 1), if it causes municipal decision-makers to actually perceive greater accountability pressure (Hypothesis 3), and/or if it causes them to become more aware of performance shortfalls (Hypothesis 4). Even then, however, effects on municipal performance outcomes could be limited by hard constraints on municipal government performance that make it impossible for municipalities to fix their performance shortfalls, for example due to a lack of funding, competence or administrative autonomy, and especially due to the ongoing armed conflict in Burkina Faso, which affects municipalities in many ways. Thus, positive effects with regard to Hypotheses 3 or 4 do not automatically imply increases in municipal performance, but could enable such improvements, if and where the right conditions are in place.

From a policy perspective, measurable impacts on municipal performance are the ultimate objective. However, a lack of positive impacts of the CBO incentive scheme on municipal performance should not per se preclude the 3PPP further consideration, especially if 3PPP shows promise in stimulating CBO involvement in municipal governance or increasing the extent to which decision-makers feel they are being held accountable for municipal performance. If those effects were to be confirmed, but fail to translate into improved municipal performance, the next logical step would

be to investigate if complementary interventions can remove the constraints that prevent greater CBO involvement in municipal governance or greater accountability pressures on local decision-makers from precipitating in improved municipal performance. Under certain conditions, it may therefore be advisable to continue testing the approach as part of a more encompassing bundle of interventions, even if the approach on its own shows little or no effect on municipal performance outcomes.

If 3PPP does cause measurable improvements in municipal performance, even under the exceptionally adverse conditions of this study (armed conflict coinciding with a period of extreme societal and political instability and the outbreak of the COVID-19 pandemic), it would be advisable to further assess its potential under circumstances that are potentially more conducive to social mobilization and direct interaction between CBOs and local decision makers, for example in other countries or during periods of greater stability. Moreover, prior to adopting 3PPP as a general policy recommendation, it would be important to assess longer-term effects on CBO involvement and municipal government performance after the 3PPP incentive has been discontinued, to ensure that short-term benefits of the approach are not outweighed by any unintended and yet unknown consequences of withdrawing the incentive.

4 Blind Analysis Methodology

To ensure the credibility and transparency of our study, we conducted a fully results-blind analysis. In a results-blind analysis, we work with blinded datasets in which all variables indicating treatment assignment or treatment status have been removed, masked, or replaced with simulated (randomly permuted) treatment identifiers. The original, unblinded dataset is encrypted and safeguarded by separate team members, who are not authors of this study, and is released to the authors only after completion of the results-blind analysis.

Blind analysis is a method to enhance research transparency and to minimize experimenter bias, as well as distortionary researcher incentives and effects of cognitive biases that could otherwise influence the interpretation of results, the allocation of effort, the choice of methods by the investigators during the process of data analysis. Blind analysis differs from a pre-registered analysis plan in that it allows researchers to take into consideration the actual outcome data and information on the actual implementation of the experiment to select the final analytical framework, data cleaning and estimation strategies. Yet, it retains the theoretical advantage of pre-registered analysis plans – committing to analytical strategies prior to knowledge of the results.

While blind analysis is common some fields of research, especially experimental physics (Klein & Roodman, 2005; MacCoun & Perlmutter, 2015), its adoption in the social sciences is rare (Humphreys et al., 2013; Lierl and Holmlund, 2019). Yet, we believe that for experiments that involve substantial uncertainty regarding implementation experiences and data quality, blind analysis offers superior flexibility, rigor and transparency compared to the more widely adopted practice of committing to pre-registered analysis plans prior to study implementation. Blind analysis not only encourages researchers to select statistical methods that are optimally suited for the actually available data, but it also ensures that the many judgments that researchers make during the process of data analysis (even when implementing a pre-specified analysis) are unaffected by prior knowledge of the results. As a consequence, a blind analysis allows for a much more comprehensive, detailed and rigorous specification of analytical methods than the typical pre-registered analysis plan, which is then often deviated ex-post from or only partially feasible.

To implement the results-blind analysis, we established the following procedures for data collection, data management and analysis.

4.1 Blinded datasets

During the data collection phase, all post-treatment data on CBOs and municipal decision-makers was handled by an outside organization, Innovations for Poverty Action (IPA). IPA was contracted to organize, manage and supervise the

endline data collection and to provide real-time data quality control. All data processing staff at IPA were specifically instructed to not to share any data with the authors that could compromise the results-blind analysis, as well as to refrain from conducting any analysis with the data. IPA also conducted a first pass of data cleaning to ensure the integrity of the dataset, which involved removing duplicate entries, identifying missing submissions, ensuring that respondent identifiers merged with the sampling frame, labeling the dataset and correcting obvious data entry mistakes. These data cleaning tasks did not require judgments by the authors and were therefore handled independently by IPA.

To generate blinded datasets, all variables that could indicate treatment assignment or treatment status were removed and the order of the observations in the dataset was randomly permuted. The removed variables included location identifiers, survey questions that were conditional on treatment assignment, as well as any other questions that could reveal treatment status. These variables were saved in a separate dataset, which was used to analyze treatment compliance, attrition and response rates and to calculate sampling weights separately from the outcomes of interest. Since identifiers were removed and the datasets were re-ordered at random, it was not possible to link the two datasets.

The blinding of the endline datasets was carried out by IPA and the blinded datasets were transmitted to the authors along with an encrypted and password-protected package containing the unblinded data, raw data and data cleaning code. This unblinded data package was safeguarded by a team of research assistants at the authors' institutions. IPA shared the password to access the unblinded data was shared only with the research assistants in charge of managing the unblinded data, not with the authors. That way, research assistants could be relied on to verify the completeness of data deliverables and to complete tasks that required access to the unblinded or raw data (such as merging baseline with endline data and then re-generating updated blinded datasets for analysis), without compromising the results-blind process. The research assistants, in turn, were asked to strictly refrain from conducting any analyses with the unblinded data of their own.

Outcome data on municipal performance was available in unblinded format, but to preserve the integrity of the of the results-blind analysis, we refrained from conducting any non-blinded analysis of the municipal performance data and used simulated treatment identifiers instead of merging the dataset with the actual treatment identifiers. We also requested implementation partners and others with access to the municipality-level treatment assignment to refrain from conducting any analyses of their own.

4.2 Blind analysis process

The blind analysis process consisted of four steps: (1) data cleaning and construction of outcome indicators from the available data, (2) selection of outcome indicators and aggregation strategy for hypothesis tests, (3) estimation strategy for treatment effects and definition of hypothesis tests, (4) presentation of blinded results and incorporation of peer feedback. At times, these steps were used iteratively to refine the data analysis strategy.

Our research hypotheses and the experimental design were first defined in a concept note that was externally reviewed in 2013. Following the proof-of-concept study in 2016, the research design was finalized and the primary research questions and outcomes of interest (H1, H3 and H5) were documented in an updated concept note in 2018, prior to implementation of the intervention.

During the data cleaning stage, we used the blinded data to construct outcome indicators and assess their suitability and data quality. This typically involved visualizing the distribution of the data, evaluating response rates and the plausibility of answers and judging whether the available data was plausibly measuring the concepts of interest. At this stage, various indicators were discarded for reasons such as evident mismeasurement and high non-response rates. During this process, blind analysis allowed for more flexibility than a pre-registered analysis plan, which would have forced us to continue considering outcome indicators that poorly measured the underlying concepts.

The selection of outcome indicators for each research hypothesis involved checking if the outcome indicators exhibited

sufficient variation to capture potential effects of the experimental intervention and if indicators measuring the same or related concepts correlated in the expected ways. At this stage, several variables were excluded from further consideration, because they exhibited little to no meaningful variation. Furthermore, we used the blinded data to select the most appropriate methods by which we aggregated multiple variables into a single hypothesis test (e.g. index construction, z-score averaging, etc.) or corrected for multiple comparisons. Blind analysis ensured that these judgments were not influenced by how they affect the ultimate treatment effect estimates and hypothesis tests.

Determining the estimation strategy for the outcomes of interests and hypothesis tests involved the choice of statistical models appropriate for the data, as well as judgments about the possibility of increasing precision through covariate adjustments. Here, blind analysis ensured that our choices of statistical methods were informed by the data, without being consciously or unconsciously driven by how they might influence the results of hypothesis tests.

We presented our results-blind analysis at several venues to solicit peer feedback from researchers and development practitioners: At the German Institute for Global and Area Studies (GIGA), the European Workshop in Empirical Political Science, the Marburg Centre for Institutional Economics Seminar and the World Bank's DIME Lightning Seminar. We incorporated the peer feedback over the course of multiple revisions of the blind analysis and solicited seminar participants' expectations about the actual results. Upon conclusion of the results-blind analysis, a written results-blind analysis report and data analysis code are uploaded to the American Economic Association's (AEA) registry of randomized controlled trials. The same data analysis code is then re-run with the unblinded data and the results-blind report is updated with the actual results and uploaded to the registry.

5 Treatment Assignment, Compliance and Attrition

Figure 3: Experimental Design

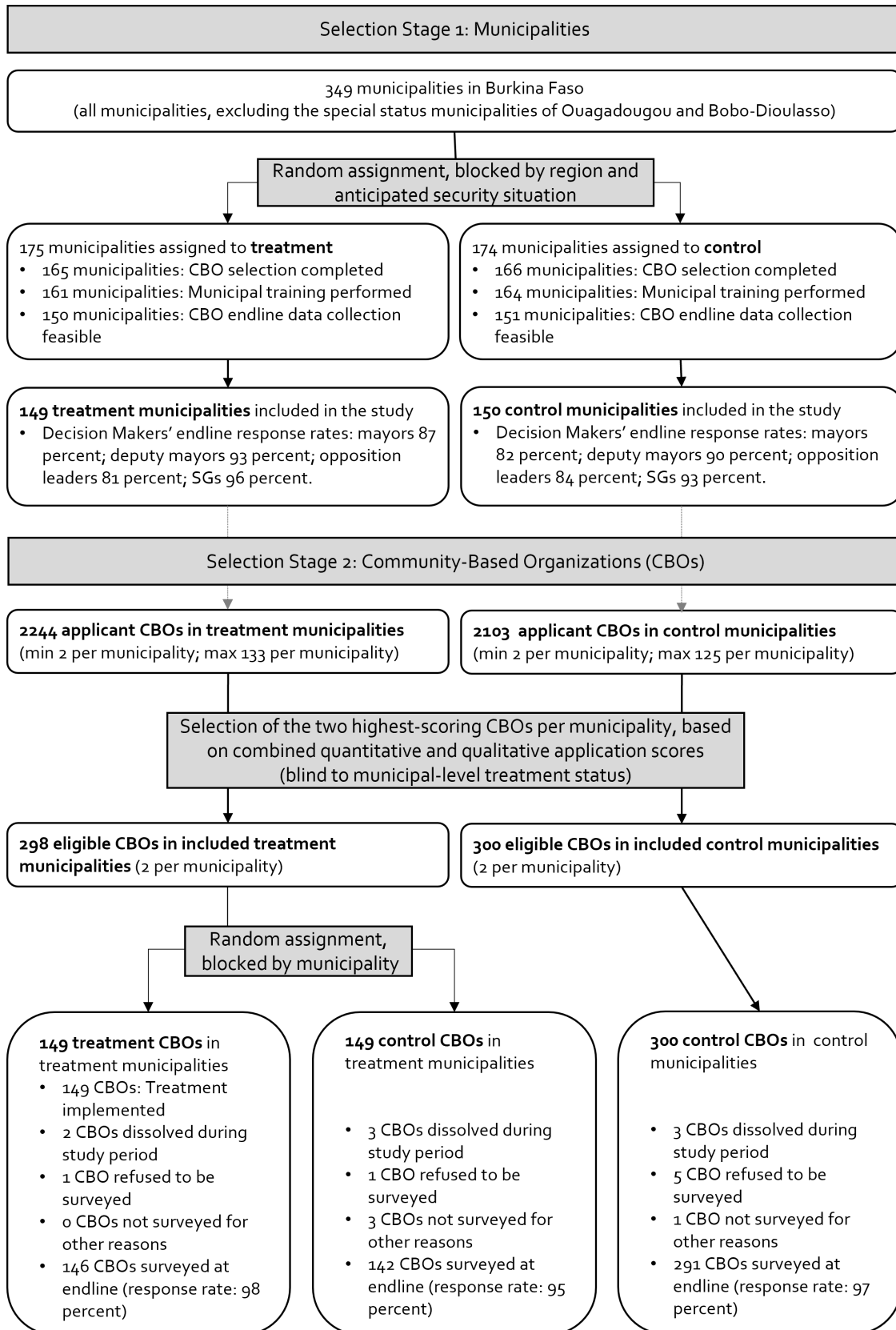


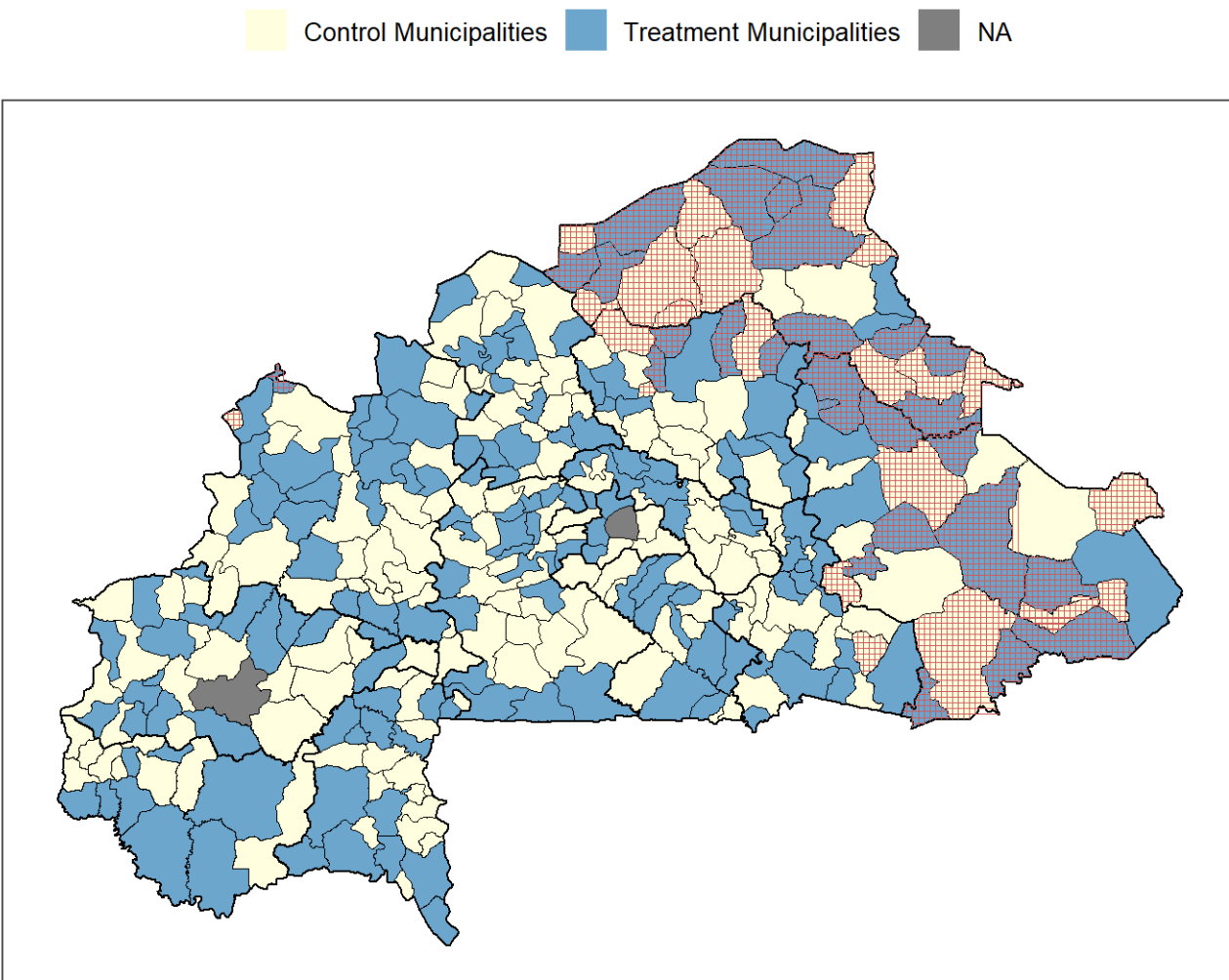
Figure 3 visualizes the treatment assignment, compliance and attrition at the municipality and CBO levels. Each step is explained in further detail below.

5.1 Treatment assignment and compliance at the municipality level

Originally, the experiment was designed to cover all 349 municipalities of Burkina Faso outside the two main urban centers of Ouagadougou and Bobo-Dioulasso. Of these 349 communes, 175 were randomly assigned to treatment (CBO incentive scheme) and 174 to control. Treatment assignment at the municipal level was block-randomized by region and anticipated security situation (which was classified in three broad categories: safe, difficult, inaccessible), in order to minimize the risk of differential attrition across treatment and control groups.

Due to the volatile security situation, there was attrition at the municipality-level even prior to treatment launch. A public call for expressions of interest by CBOs could be launched in only 335 out of 349 municipalities (168 treatment and 167 control). Out of these 335 municipalities, a sufficient number of applications were received in 334 municipalities (167 treatment and 167 control). Additionally, due to operational delays on the part of the implementing partner, the expressions of interest in three municipalities (two treatment and one control) were not received in time for the selection of partner CBOs. As a result, these municipalities were dropped from the study. The remaining 331 municipalities (165 treatment and 166 control) thus defined the initial coverage of the experiment. However, as the security situation deteriorated over the course of the intervention period, further municipalities had to be excluded, either because it was not feasible to disseminate municipal performance scorecards in those communes, or because it would not have been possible to collect endline data in those communes. As a result, we were left with 299 communes (149 treatment and 150 control) in which the experiment was feasible. In these 299 communes, our implementing partners complied with the assigned treatment status.

Figure 4: Map of treatment and control municipalities. Shaded areas indicate municipalities that were excluded from the study due to security concerns.



5.2 Treatment assignment at the CBO level

Among the two CBOs whose applications ranked highest within their municipality, one was randomly assigned to the treatment group and the other to the internal control group. In control municipalities, the two highest-scoring CBOs were both selected into to the external control group. In total, 332 CBOs in control municipalities were assigned to the external control group, while in treatment municipalities, 165 CBOs were assigned to the internal control group and 165 to the treatment group. Among the shortlisted CBOs, it turned out that three CBOs had submitted applications in two municipalities instead of one and ended up being ranked among the top two CBOs in both of the respective municipalities. These double-applications affected a total of four municipalities. The issue was resolved by dropping each double-applicant from the shortlist in one of the municipalities in which they applied and replacing them with the next-highest ranked CBO in that municipality, while retaining them on the list of eligible CBOs in their other municipality.

5.3 Attrition and non-response at the CBO level

Response rates in the CBO endline survey were high and homogeneous across treatment conditions, ranging between 95 and 98 percent within the 299 municipalities included in the study. A total of eight CBOs had dissolved during the study period and were therefore not surveyed at endline. These cases are distributed about proportionally among the different treatment conditions.

Group-based surveys with were conducted with up to 9 participating CBO members, including members of the CBO’s leadership as well as ordinary members. These individual survey participants were also interviewed separately in all of the surveyed CBOs. For two CBOs in one municipalities, a technical malfunction of data collection equipment led to data loss that affected these individual-level questionnaires.

5.4 Attrition and non-response at the level of municipal decision makers

Data collection at the level of municipal decision makers consisted of individual interviews with majors, deputy mayors, secretary generals of the municipality and local opposition leaders. These were carried out in a total of 324 municipalities, including some that were dropped from this study, because the survey contained modules for other purposes as well. In this study, we only consider respondents from the 299 municipalities where the experiment was feasible. Within the 299 study municipalities, response rates by municipal decision-makers were high and similar across treatment and control municipalities (over 80 percent for mayors and opposition leaders, over 90 percent for deputy mayors and SGs). In two municipalities, a subset of the data from decision-maker interviews had to be excluded from the analysis, because the surveyors accidentally asked them about the wrong CBOs (replacements, as mentioned above, were not properly taken into account during the programming of the survey tools).

5.5 Balance tests

In the 299 municipalities where the experiment and endline data collection were feasible, treatment and control communes are well-balanced with respect to baseline outcomes, geographic characteristics of municipalities and conflict exposure (Table 2). We also find no statistically or substantially significant differences in the response rates of CBOs and municipal decision-makers in the endline survey. Finally, the CBO selection process (which was blind to the municipality’s treatment status) had very similar outcomes in treatment and control municipalities, both in terms of applicant numbers and in terms of applicant quality at every step of the selection process (initial evaluation, shortlisting, and final selection).

In the further analysis, we treat security-related attrition as independent of treatment status, because (1) we see no plausible reason why security-related attrition at the municipal level could have been in any way endogenous to a municipality’s assigned treatment status, (2) in line with this expectation, treatment and control groups were almost equally affected by attrition (which is aided by the fact that we block-randomized by anticipated security situation), and (3) treatment and control municipalities continue to be well-balanced on all relevant covariates. We therefore restrict our analysis to the 299 municipalities where the experiment and endline data collection were feasible.

Since response rates are both very high (about 97 percent at the CBO level and between 82 and 96 percent at the municipal decision-maker level) and well-balanced across treatment and control municipalities, we also treat non-response as independent of treatment status and do not regard it as a threat to either the internal or external validity of the experiment.

	<i>Sample Mean/Proportion</i>		P-value (no difference)
	Treatment municipalities	Control municipalities	
<i>CBO selection process</i>			
Number of CBO applications	15.1	14	0.56
Quantitative score of CBO applications	36.2	35.8	0.194
Total (quantitative + qualitative) score of shortlisted CBOs	56.8	56.1	0.379
Total (quantitative + qualitative) score of selected CBOs	56.8	56.1	1
<i>Survey response rates</i>			
Response rate CBO survey	0.42	0.41	0.746
Response rate mayors	0.872	0.82	0.272
Response rate opposition leaders	0.812	0.84	0.628
Response rate deputy mayors	0.926	0.9	0.55
Response rate SGs	0.96	0.927	0.325
<i>Baseline municipal performance</i>			
2018 SUPERMUN performance score	117	115	0.345
2018 SUPERMUN institutional capacity score	45.5	44.1	0.23
2018 SUPERMUN service delivery score	72.5	73.5	0.442
<i>Baseline conflict exposure</i>			
Fraction of municipalities with 'Impossible' baseline security situation	0.00671	0.00667	1
Fraction of municipalities with 'Difficult' baseline security situation	0.0201	0.0267	1
Conflict fatalities in 2018 (ACLED)	26	21	0.69
<i>Municipality Characteristics</i>			
Fraction of municipalities classified as 'urban'	0.141	0.141	1
Population estimate in 2018	38600	40300	0.552
Number of municipalities	149	150	.

Table 2: Balance tests - Municipalities included in the analysis

6 Estimation Strategy and Results-Blind Analysis

6.1 Hypothesis 1: A financial stake in municipal government performance increases CBOs' involvement in municipal governance.

6.1.1 Outcomes of interest and data sources for Hypothesis 1

To test Hypothesis 1, we look at four aggregate outcome indicators that are each constructed from multiple individual outcomes indicators: (1) The breadth of CBOs' involvement in municipal affairs, (2) the intensity of CBOs' involvement in municipal affairs, (3) municipal decision makers' knowledge of an interaction with CBOs, (4) municipal decision makers' perception of CBOs' influence in municipal governance.

Table 3: Hypothesis 1 - Outcome Indicators

Outcome Indicator	Components	Question Number	Unit of observation
Breadth of CBO engagement in	Has met: Municipal Council	INT1	CBO
	Has met: Cadre de Concertation.	INT3	

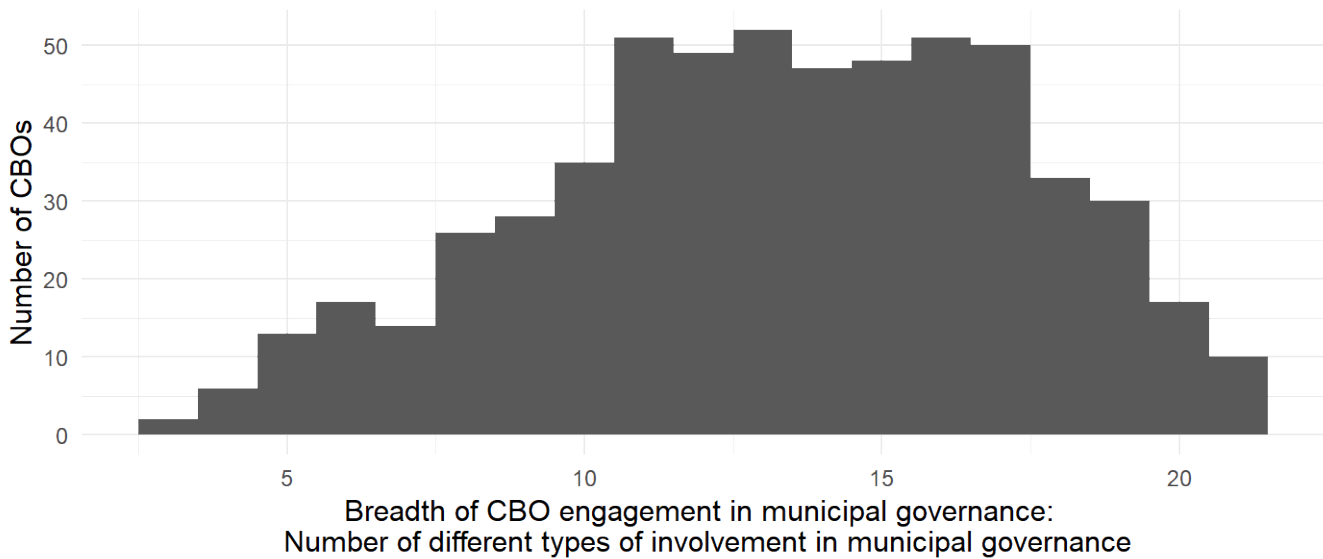
Outcome Indicator	Components	Question Number	Unit of observation
municipal governance	Has met: Secretary General.	INT6	
	Has met: Municipal Administration.	INT10	
	Has met: Mayor.	INT14	
	Has met: Councilor.	INT18	
	Has met: Prefecture.	INT23	
	Has met: Health Authorities.	INT24	
	Has met: Education Authorities.	INT25	
	Has met: Water Authorities.	INT26	
	Has met: Village chiefs / CVD.	INT27	
	Has Collected: data on municipal services.	RES1	
	Has Collected: opinions about muni. governance, or local public services .	RES3	
	Has interviewed service providers.	RES5	
	Has presented at village meetings.	ORG1	
	Has collaborated with other CBOs or equivalent.	ORG4	
	Has organized: multistakeholder meetings.	ORG8	
	Has organized: meetings with CBOs or equivalent.	ORG5	
Has created a public whatsapp group or another kind of social media forum.	ORG9		
Has organized an outreach or information campaign about an issue of public interest in muni..	PRES1		
Intensity of engagement in municipal governance	Number of meetings with elected officials: Mayor.	INT14	CBO
	Number of meetings with elected officials: Municipal Council.	INT1	
	Number of meetings with elected officials: Councilor.	INT18	
	Number of meetings with elected officials: Cadre de Concertation.	INT3	
	Number of meetings with unelected officials: Secretary General.	INT6	
	Number of meetings with unelected officials: Municipal Administration.	INT10	
	Number of meetings with service providers: health.	INT24	
	Number of meetings with service providers: education.	INT25	
	Number of meetings with service providers: water.	INT26	
	Number of organized multi-stakeholder meetings.	ORG8	
	Number of CBOs or equivalent: collaborated with.	ORG4	
	Number of villages in which the presented.	ORG2	
Number of organized outreach or information campaign.	PRES1		
Decision makers' knowledge of and interaction with CBOs	0: The decision maker: doesn't know the CBO	B101	Decision makers × CBO
	1: The decision maker: has heard about the CBO and. . .	B101	
	2: . . . knows at least one member and. . .	B102	

Outcome Indicator	Components	Question Number	Unit of observation
	3: ... knows that CBO has approached the municipal administration or the council and...	B103	
	4: ... has been approached by the CBO.	B104	

Breadth of CBO involvement in municipal governance

To measure the breadth of CBOs' involvement in municipal governance, we count the number of different activities a CBO has engaged in during the study period, out of a list of 20 different indicators of involvement in municipal affairs. Although this outcome variable is censored at zero and 20, its distribution suggests that floor and ceiling effects are negligible (Figure 5).

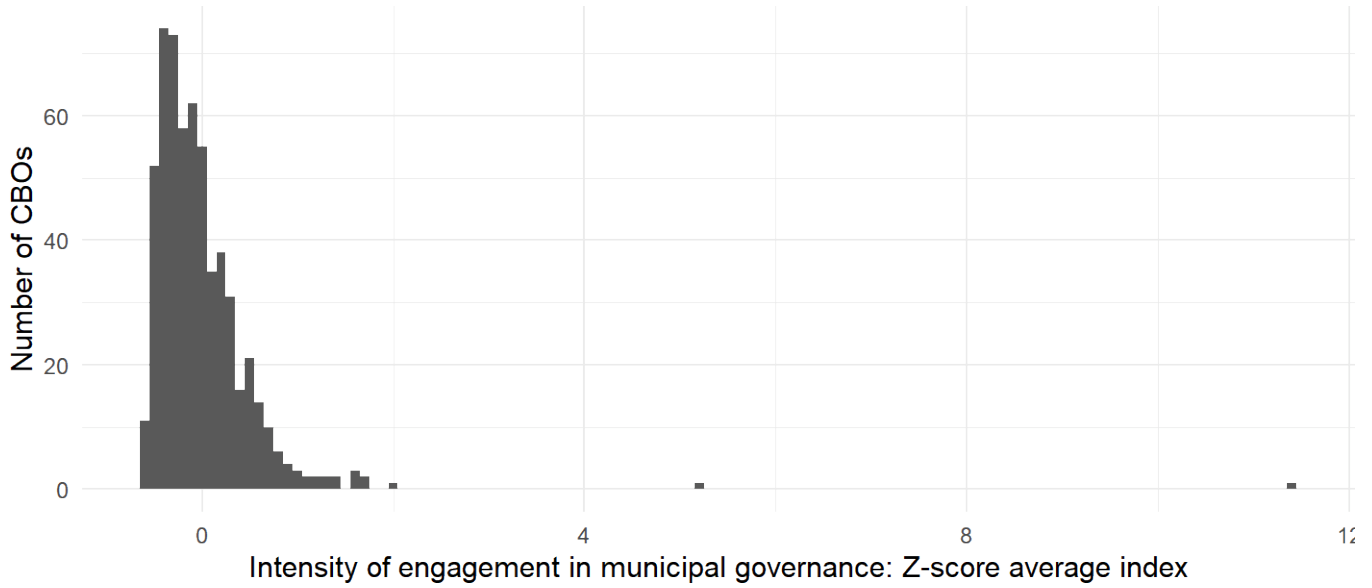
Figure 5: Breadth of CBO involvement in municipal governance (pooled data, treatment and control CBOs)



Intensity of CBO involvement in municipal governance

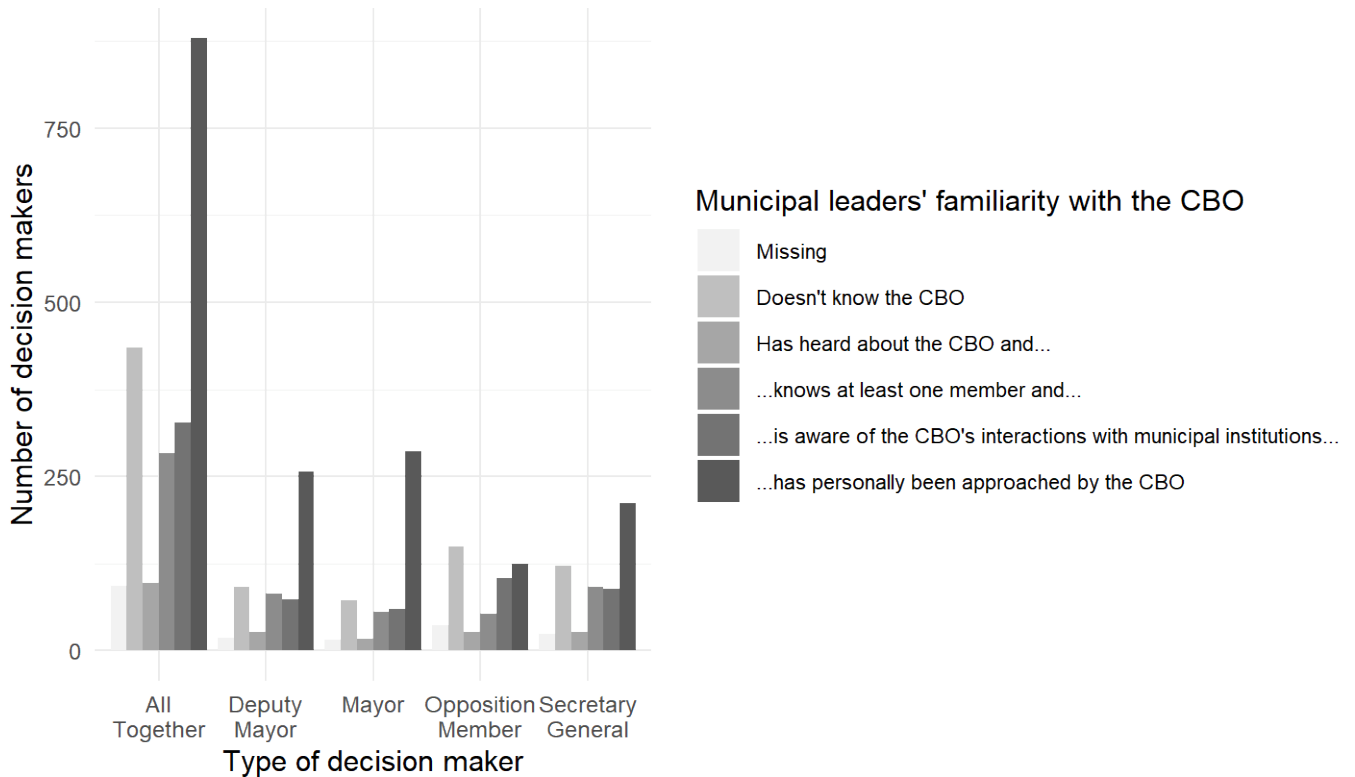
To measure the intensity of the CBOs' engagement in municipal governance, we build a summary index using the average of several standardized indicators (z-score average) that measure the frequency of specific actions by the CBO within a specified time frame. The actions include, inter alia, meetings with the mayor and other decision-makers, meetings with service providers, presentations at village meetings, information campaigns, etc. The time frame depends on the question and refers to either a one year period or the years 2019 and 2020 combined.

Figure 6: Intensity of CBO involvement in municipal governance (pooled data, treatment and control CBOs)



Municipal decision makers' knowledge of and interaction with CBOs

Figure 7: Municipal decision makers' knowledge of and interaction with CBOs (pooled data, treatment and control CBOs)



To quantify top-level municipal decision makers' (mayor, deputy mayor, secretary general and local opposition leader) awareness of and interaction with CBOs, we construct an outcome variable that sum up four binary indicators of whether:

- the decision maker has heard about the CBO,

- the decision maker knows any member of the CBO,
- to the knowledge of the decision maker, the CBO has ever actively approached the municipal administration or the municipal council about an issue of municipal governance,
- the decision maker, in his official role, has personally ever been approached by the CBO about a problem of municipal governance.

The minimum value of 0 indicates that the decision maker has never heard of the corresponding CBO. The maximum value of 4 indicates that the responding decision maker has personally been approached by the CBO about a problem of municipal governance.

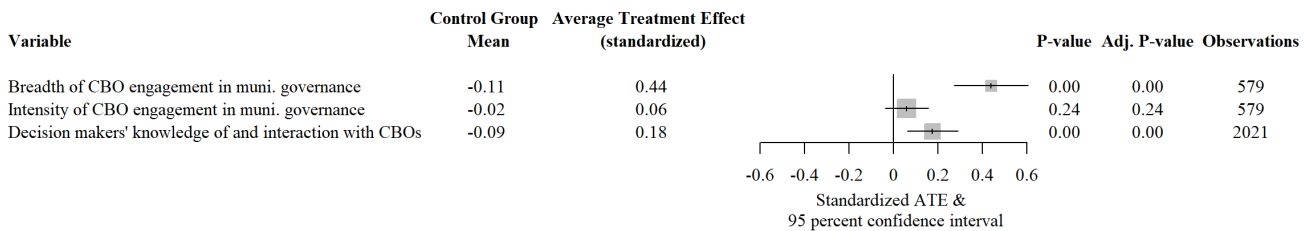
6.1.2 Estimation strategy for Hypothesis 1

For the breadth and intensity of CBO involvement in municipal governance, as well as municipal decision makers' knowledge of and interaction with the CBO, we estimate average treatment effects via OLS regression, including fixed effects for the randomization blocks (the combination of region and anticipated security situation).

For outcomes at the level of decision makers, we adjust for clustering of errors at the CBO level. Additionally, when pooling the data across different types of decision makers (mayors, deputy mayors, opposition leaders and SGs), we weight each observation by the inverse of the number of decision makers interviewed in the municipality, so that equal weight is assigned to all CBOs, even if data is missing for some decision makers in some municipalities.

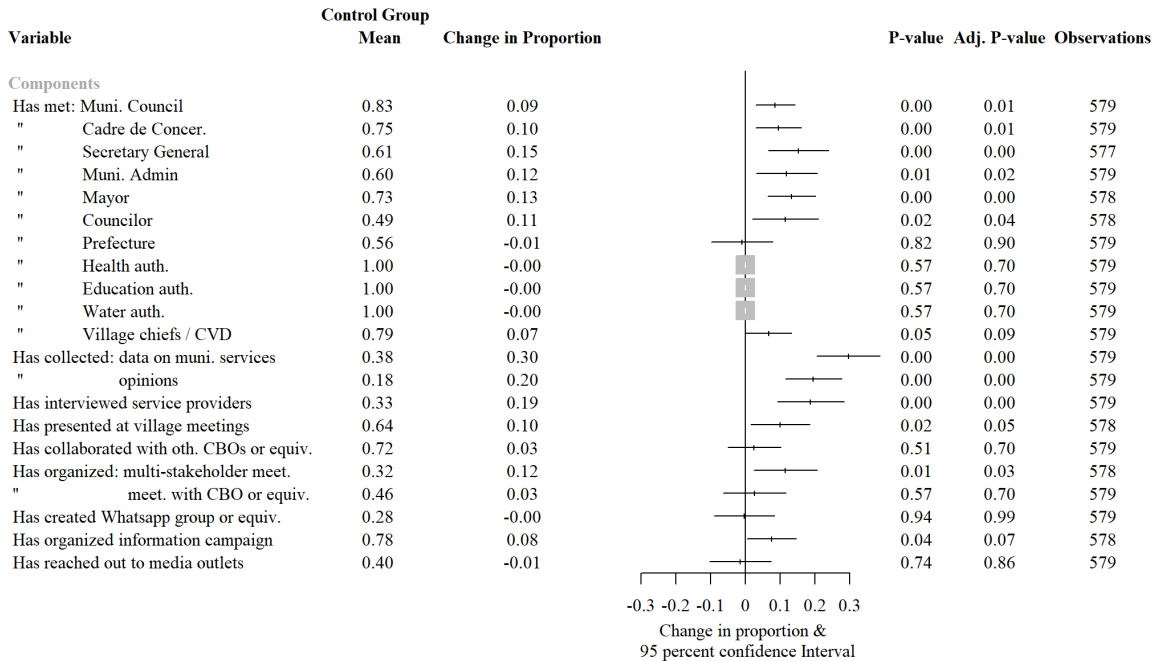
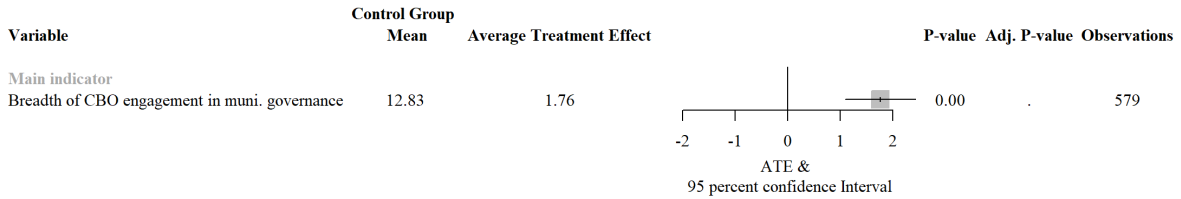
6.1.3 Unblinded results for Hypothesis 1

Figure 8: Hypothesis 1 - Aggregate outcome variables



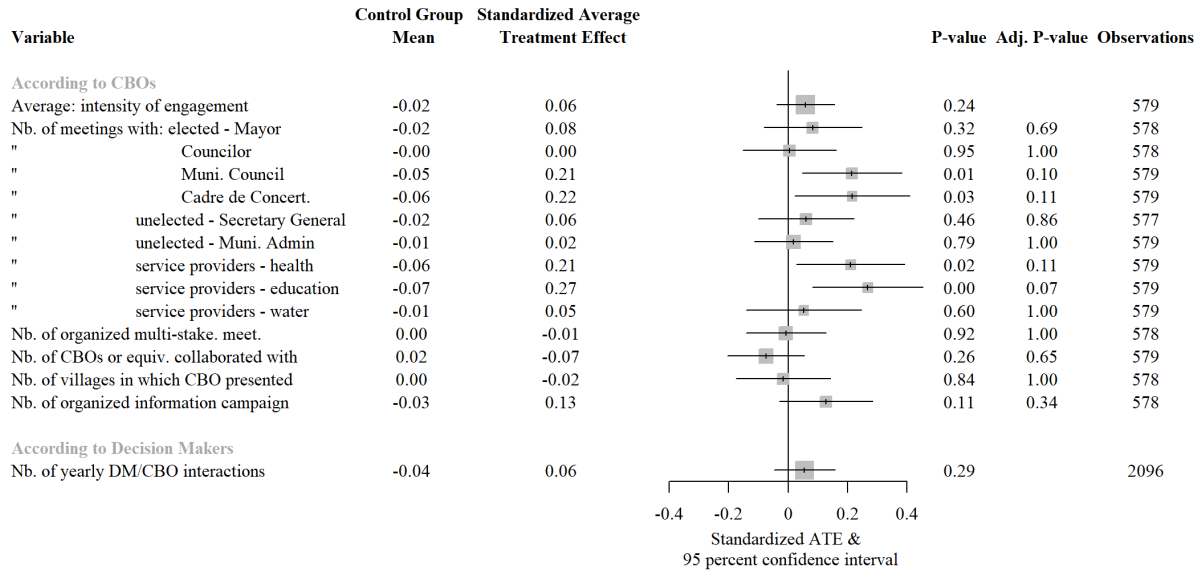
Notes: OLS coefficient estimates and 95 percent confidence intervals. All specifications include fixed effects for the randomization strata (region and anticipated security situation). The estimations for breadth and intensity of CBO engagement are performed at the CBO level. The one for the decision makers' knowledge of and interaction with CBOs is performed on a dataset where observations represent every decision maker / CBO combinations available in the decision maker dataset. It also adjusts for clustering at the CBO level (640 clusters) and controls for the type of decision maker. The estimation for the decision makers' perception of CBOs' influence is run at the decision maker level and is restricted to treated municipalities. Weighting equal to the inverse of the number of decision makers interviewed in the municipality is applied to each observations of the last two estimations. Standardized control group means are displayed. P-value adjustment is performed to control the false discovery rate (FDR) following Benjamini & Hochberg (1995).

Figure 9: Treatment effects on breadth of CBO involvement in municipal governance



Notes: OLS coefficient estimates and 95 percent confidence intervals. All specifications include fixed effects for the randomization strata (region and anticipated security situation). All components of the main indicator are binary variables. P-value adjustment is performed to control the false discovery rate (FDR) following Benjamini & Hochberg (1995).

Figure 10: Treatment effects on intensity of engagement in municipal governance



Notes: OLS coefficient estimates and 95 percent confidence intervals. All specifications include fixed effects for the randomization strata (region and anticipated security situation). Estimations in the "According to CBOs" subgroup are run at the CBO level. The estimation in the "According to Decision Makers" subgroup is run on a dataset where observations represent every decision maker / CBO combinations available in the decision maker dataset. It also adjusts for clustering at the CBO level (640 clusters) and controls for the type of decision maker. Weighting equal to the inverse of the number of decision makers interviewed in the municipality is applied to each observations of the last two estimations. P-value adjustment is performed to control the false discovery rate (FDR) following Benjamini & Hochberg (1995).

6.1.4 Alternative outcome indicators and estimation strategies considered during blind analysis

During blind analysis, we explored several alternative ways of grouping and aggregating indicators of CBO involvement in municipal affairs. At the CBO level, we decided to separately consider the breadth and intensity of CBO involvement, because there could be effects along both the extensive and intensive margins of different types of involvement in municipal affairs. CBOs might take up new activities to influence municipal performance, or they might dedicate more time to specific activities they have already previously engaged in.

An alternative approach would have been to group indicators into several distinct categories of activism, such as meetings with decision-makers, information gathering, community mobilization and public pressure, problem-solving with service providers, etc. Although our data collection instruments were structured that way, we opted not to break down CBOs' activism into different categories, because CBOs are likely to differ in their approaches and strategies of influence, so that average effects in any one category might have been too small to detect with the available sample size, especially after adjusting for multiple testing.

To aggregate the indicators of intensity of CBO involvement in municipal affairs, we considered using an index based on principal components (PCA) or on inverse covariance weighting (ICW). PCA clustered the indicators into at least three distinct and substantively relevant dimensions of variation, which would have left us with the choice of having either three aggregate variables or discarding/underweighting relevant information. ICW produced negative weights for some highly correlated but substantively relevant indicators, which would have made it difficult to interpret the index. We therefore opted for z-score averaging, which gives equal weight to all indicators.

Finally, we opted to disregard a survey module on the perceived influence of CBOs. In this module, municipal decision-makers were asked to compare treatment and control CBOs along their ability to mobilize people around an issue of

<i>DV: Awareness of and interaction with CBOs</i>					
<i>0 = Has never heard about CBO ... 4 = Has personally been approached by the CBO</i>					
	(1) All	(2) Mayors	(3) Deputy Mayors	(4) Opposition Leaders	(5) Secretary Gen- eral
Control group mean	2.41	2.83	2.55	2.03	2.24
Average Treatment Effect	0.287*** 0.0906	0.232* 0.134	0.37*** 0.132	0.0501 0.158	0.411*** 0.137
Observations	2021	491	530	458	542
Clusters (CBOs)	590				

Coefficients from OLS regressions with fixed effects for randomization blocks. Column (1): Cluster-robust standard errors in parentheses. Columns (2)-(5): HC3 standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided).

Table 4: Effects on municipal decision-makers' knowledge of and interaction with CBOs

public interest, constructively participate in municipal governance, put pressure on municipal decision-makers, their overall power within the commune and their level of organization. To keep the surveys identical, these questions were asked both in the treatment and control municipalities, but the data are meaningful only in treatment municipalities, where there are treated CBOs. However, during blind analysis, it became apparent that only a subset of decision-makers even knew about the CBOs. Since such direct comparisons were only feasible if decision-makers knew both the treatment and control CBOs, the available data were not only limited to begin with, but also biased, if the treatment caused treatment CBOs to be known to more municipal decision-makers than otherwise. In that case, some previously unknown treatment CBOs would be compared to already well-known control CBOs, which would have produced a biased comparison that understates the true the effect of treatment on CBOs' perceived influence.

6.2 Hypothesis 2: 3PPP does not erode the internal functioning or collective action capacity of CBOs.

6.2.1 Outcomes of interest and data sources for Hypothesis 2

To test Hypothesis 2, we use a set of ten outcome indicators that reflect four dimensions of the CBOs' internal functioning: (1) collective action, (2) organizational capacity, (3) internal cohesion and (4) organizational capture (Table 3).

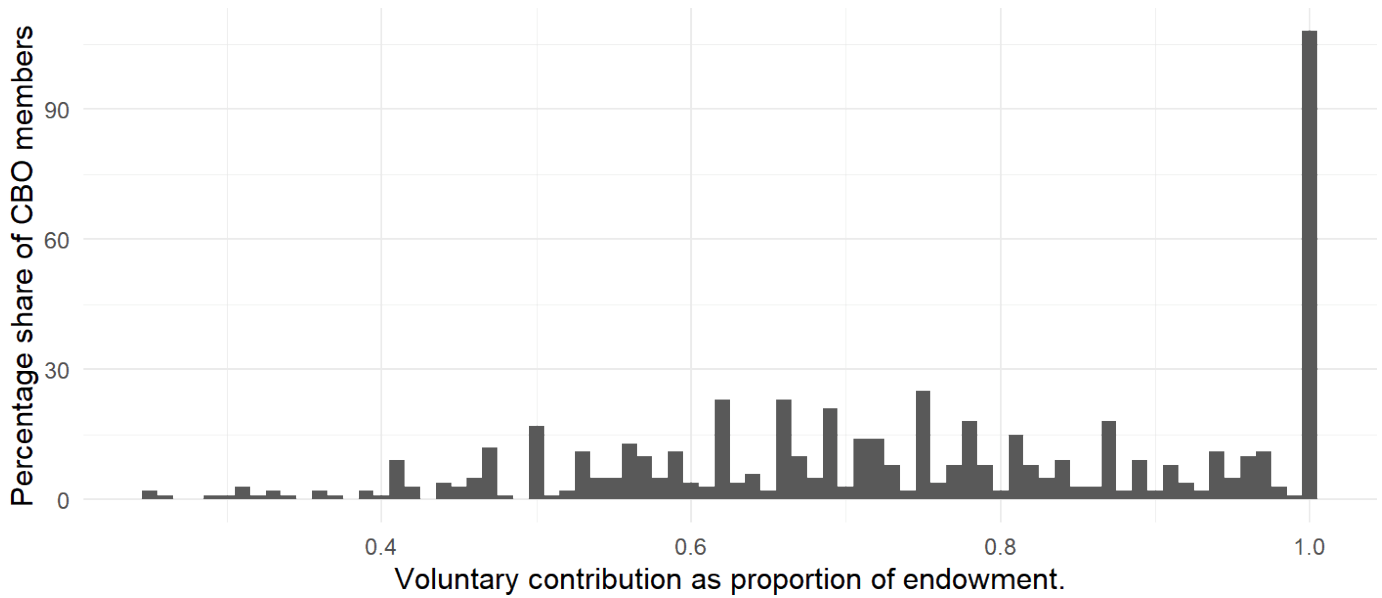
Table 5: Hypothesis 2 - Main outcome variables

Main Outcome Variables	Components	Survey item(s)	Level
Collective action capacity	Public goods game: proportion of endowment contributed	-	Individual
Organizational capacity	Number of general membership meetings held by the CBO in 2020	ACT1	CBO
	Number of executive board meetings held by the CBO in 2020	ACT2	
	CBO had a formal budget in 2020	FIN1	
	Net membership change	ACT3, ACT4, ACT4a, ACT5	
	CBO has adopted COVID-19 policy	COV2	
	CBO has seized COVID-related opportunities	COV5	
Internal cohesion	Perceived internal conflict	IND15	
	Satisfaction with the CBO leadership	IND16	
	Enthusiasm about being a member	IND17	
Organizational capture	Are the CBO's activities being abused for purposes that are not of shared interest	IND18	Individual

Collective action capacity

To measure collective action capacity among CBO members, we carried out a voluntary contributions public goods game with a group of typically eight CBO members that took part in the CBO survey. Each individual was asked to split an endowment of 2000 FCFA between themselves and a group fund in increments of 100 FCFA. Any amount allocated to the group fund was automatically doubled and the total contributions shared equally among group members. Any amount allocated to themselves would be paid directly to the respondents at the same time as the group fund was divided up. The decision was made in private by allocating money between two jars behind a screen and then recorded by the surveyor.

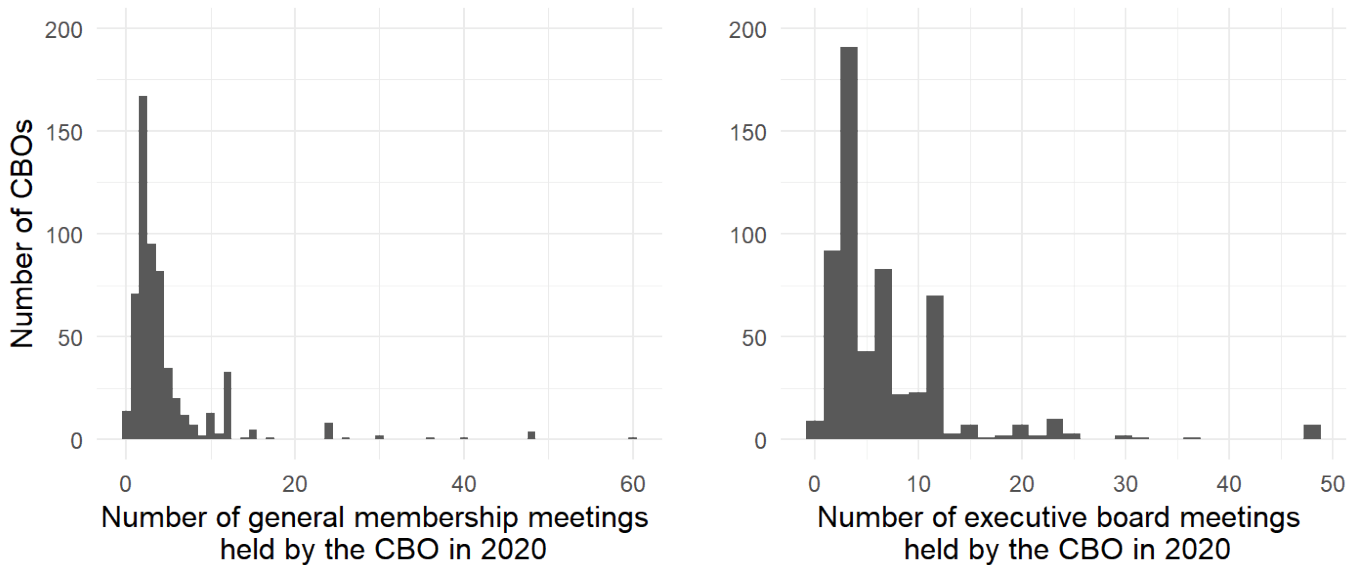
Figure 11: Public goods game: proportion of endowment contributed (pooled treatment and control CBOs)



Organizational capacity

Data on the number of general membership meetings and executive board meetings (that were not also general membership meetings) that were held in 2020 by their organization was collected through the group-based CBO survey.

Figure 12: Number of general membership and executive board meetings held by the CBO in 2020 (pooled treatment and control CBOs)

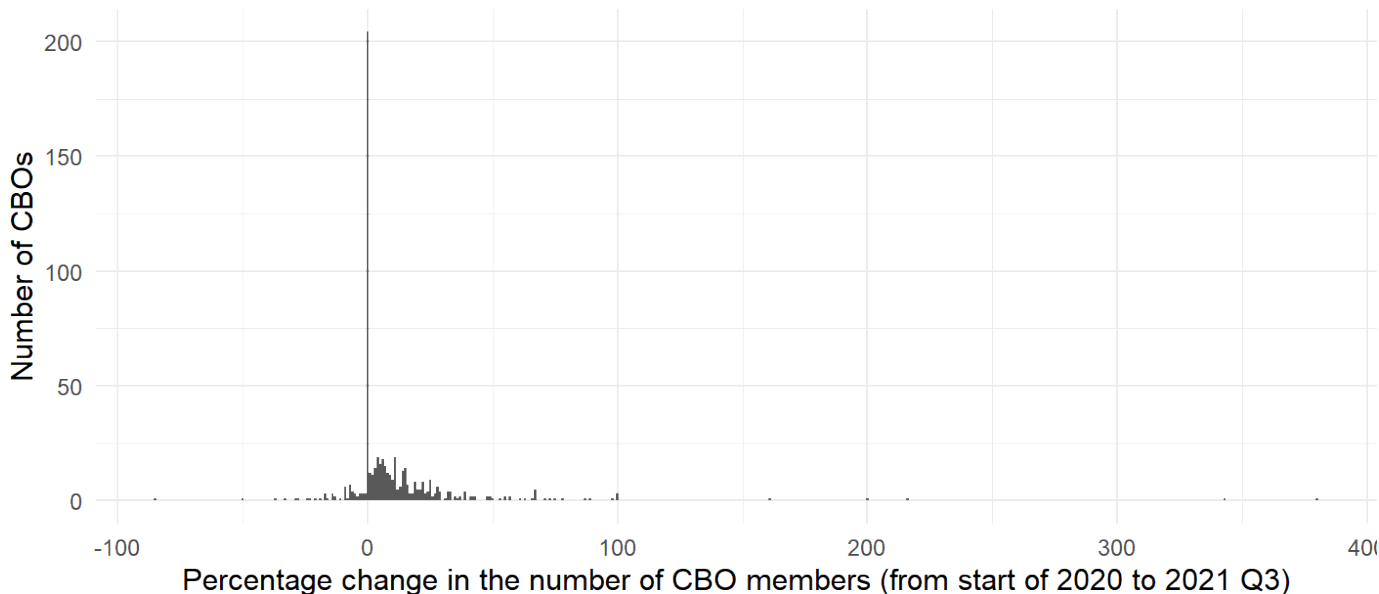


Data on whether the CBO had a formal budget in 2020 was likewise collected through the group-based CBO survey. Data is available from 588 CBOs, with 500 (approx. 85 percent) of them having a formal budget.

Data on the net membership change between 2020 and 2021 (i.e. during the intervention period) is available from the CBO survey. We calculate the net change in membership by reconstructing the number of members at the beginning of 2020 using data on (1) the number of CBO members at the time of the interview in 2021, (2) the number of members

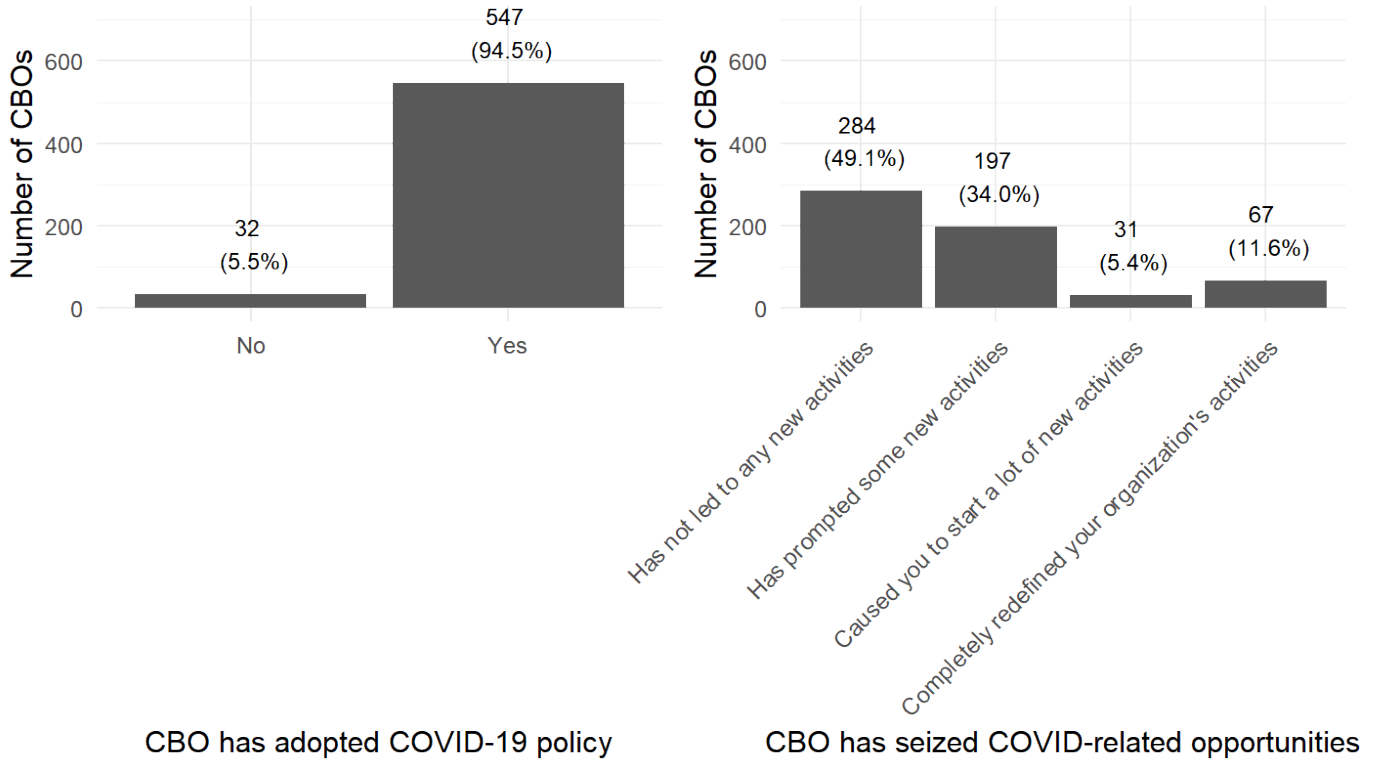
who joined since the beginning of 2020, (3) the number of members who left since the beginning of 2020. From the latter, we subtract deceased members who passed away during this time period.

Figure 13: CBOs' net membership change (pooled treatment and control CBOs)



Since the COVID-19 pandemic began shortly after intervention launch and universally affected all CBOs, it provides a useful way of measuring a CBO's adaptability to new circumstances. We focus on two indicators that reflect aspects of a CBO's organizational capacity: Whether the CBO has adopted an internal COVID-19 safety policy, and to what extent the organizations were able to quickly adapt the scope of their activities to the pandemic, by adopting new pandemic-related activities, rather than simply winding down previously ongoing activities. Also from an ethical perspective, these indicators are of interest: After an initial safety assessment, we decided to continue the CBO incentive scheme throughout the COVID-19 pandemic, but to explicitly monitor if it had any adverse consequences on the CBOs' COVID-19 resilience.

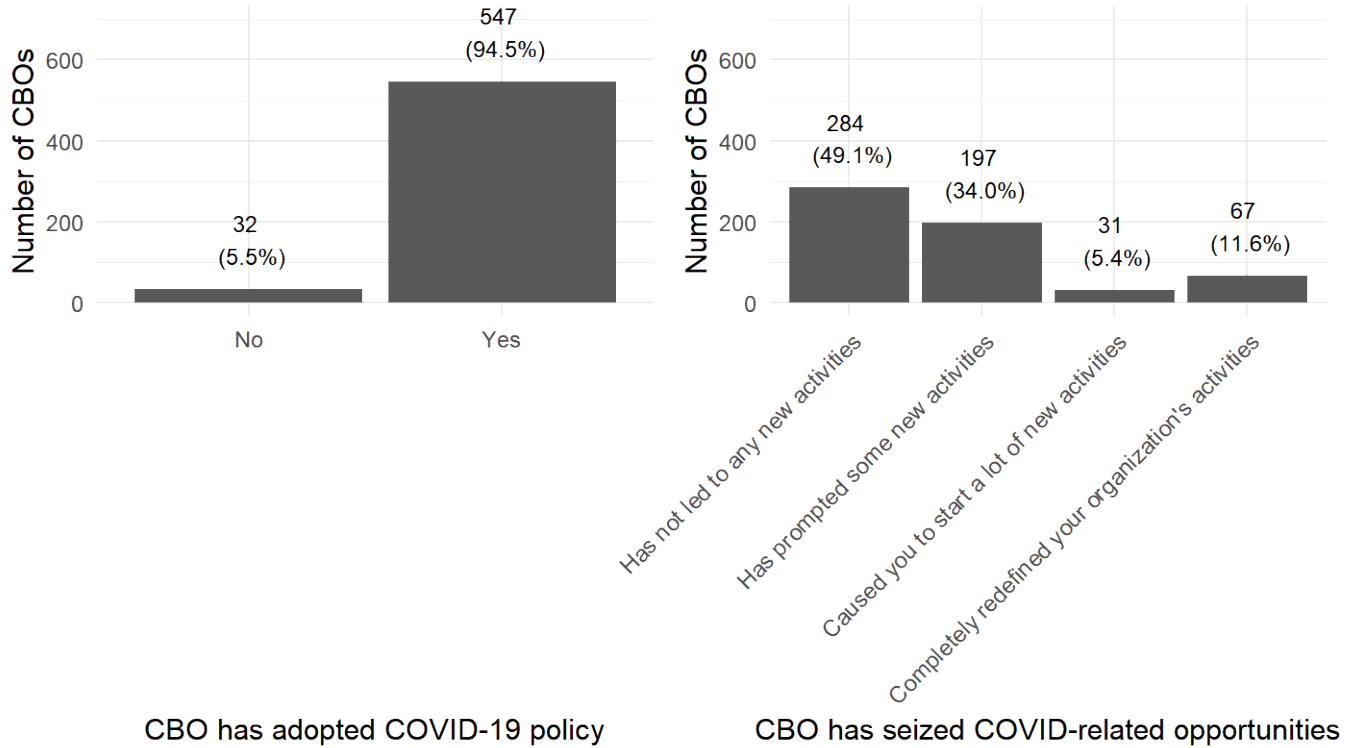
Figure 14: Covid-related organizational capacity (pooled data, treatment and control groups)



Internal cohesion

To measure the internal cohesion of CBOs, we surveyed individual members on (1) perceptions of internal conflict ("Would you say that there is often conflict among the members of [CBO name]?"), (2) satisfaction with the CBO leadership ("How satisfied are you with the leadership of the organization?"), and (3) their level of frustration or enthusiasm about being an member of the CBO, on a five-point scale from -2 to 2. ("Overall, how do you feel about your organization? Would you say that you are: (-2) Very frustrated with it, (-1) A bit frustrated with it, (0) Feeling neutral about it, (1) Feeling positive about it, (2) Feeling very enthusiastic about it?").

Figure 15: Internal cohesion indicators (pooled data, treatment and control groups)



Organizational capture

Perceptions of organizational capture are measured through a binary choice question

Would you say that your organization's activities mainly serve the COMMON INTERESTS OF ITS MEMBERS, or would you say that your organization is being ABUSED FOR PURPOSES THAT ARE NOT IN THE COMMON INTEREST of its members?

In the pooled data from treatment and control groups, 91.4 percent (4048 respondents) did not think of their organization as being captured for private interests, 8.3 percent (368 respondents) did think of their organization as being captured for private interests. 11 CBO members refused to answer.

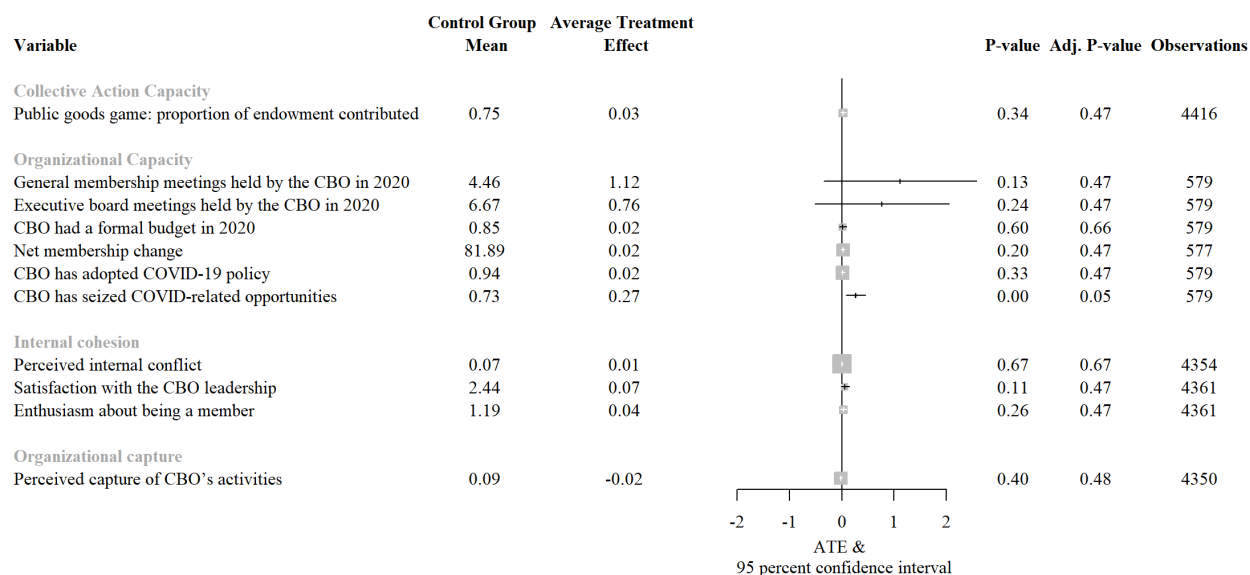
6.2.2 Estimation strategy for Hypothesis 2

We estimate average treatment effects via OLS regression, including fixed effects for the randomization strata (region and anticipated security situation). For outcome variables measured at the individual level, we adjust errors for clustering at the CBO level. These individual-level outcomes are: members' time contribution to CBO activities, perceived internal conflict, satisfaction with the CBO leadership and enthusiasm about being a member. In the case of outcomes measured at the individual level, each observation is weighted by the inverse of the number of members interviewed per CBO, so that equal weight is assigned to all CBOs, even if data are missing for some individuals in some CBOs.

Effects on voluntary contributions in the public goods game are estimated via a Tobit regression, due to a high number of individuals contributing their full endowment of 2000 FCFA. This suggests that relative to the general population, the CBO members are highly cooperative with one another, and that simple OLS estimates of treatment effects might therefore be distorted by ceiling effects.

6.2.3 Unblinded results for Hypothesis 2

Figure 16: Hypothesis 2 - Results summary



6.2.4 Alternative outcome indicators and estimation strategies considered during blind analysis

As an indicator of collective action capacity, we had planned to measure individual members' time contribution. However, we found this variable to be mismeasured: Due to an error on the part of our data collection partner, the wrong response options were displayed to this question. We therefore excluded this variable from the analysis.

Furthermore, as indicators of organizational resilience, we also collected data on CBOs' adaptation to the changing security situation. However, since CBOs were affected by insecurity to a varying extent, we did not consider this indicator sufficiently comparable across CBOs. Instead, we opted to use CBOs' COVID-19 response as an indicator of organizational resilience, because the COVID-19 outbreak affected CBOs relatively homogeneously and simultaneously during the study period.

6.3 Hypothesis 3: 3PPP causes municipal decision makers to perceive greater accountability pressure.

6.3.1 Outcomes of interest and data sources for Hypothesis 3

To test if the CBO incentive scheme causes municipal decision-makers to perceive greater accountability pressures, we focus on perceived pressure from civil society groups. We ask decision makers about the three most important actors or institutions they need to keep happy in order to maintain their influence. Our outcome of interest is whether community groups and local civil society organizations are among these three most important actors. Similarly, we ask decision makers to identify the three actors or institutions to whom they spend the most effort justifying their decisions. Our outcome of interest is whether community groups and local civil society organizations are among these three actors or institutions that require the most justification effort. Finally, decision makers are asked about the likelihood of facing civil society pressure if their municipality fails to deliver on its responsibilities without good justification. This outcome is measured on a six point probability scale ranging from 0 (impossible) to 1 (certain) in increments of 0.2. Figure 15 shows the distribution of these three outcome variables.

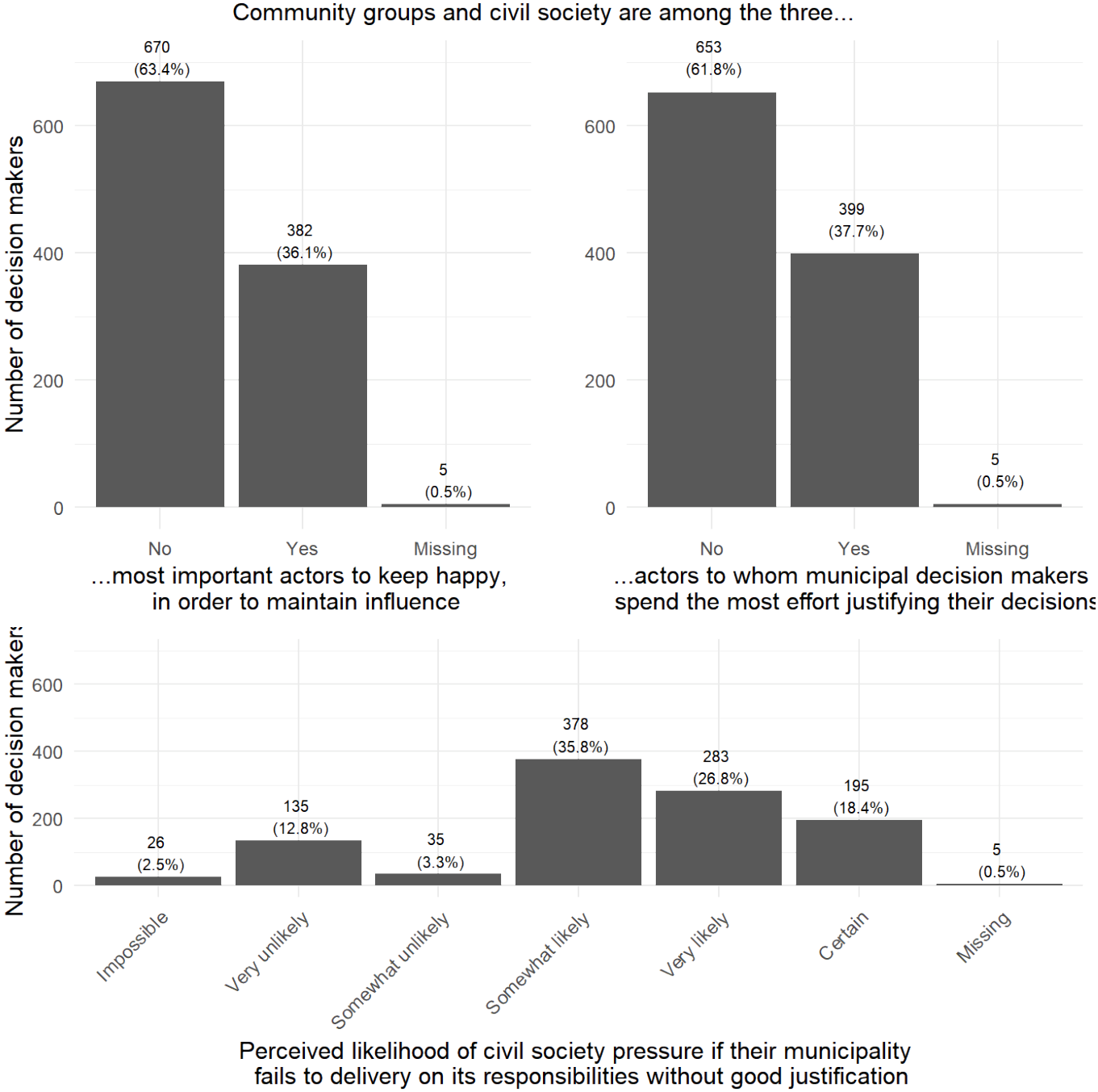
Table 6: Hypothesis 3 - Main outcome variables

Main Outcome Variables	Survey item	Unit of observation
Community groups and civil society organizations are among the three most important actors to keep happy in order to maintain influence	F101, F102, F103	Decision makers
Community groups and civil society organizations are among the three actors to whom municipal decision makers spend the most effort justifying their decisions	F104, F105, F106	Decision makers
Perceived likelihood of civil society pressure if their municipality fails to deliver on its responsibilities without good justification	G103	Decision Makers

6.3.2 Estimation strategy for Hypothesis 3

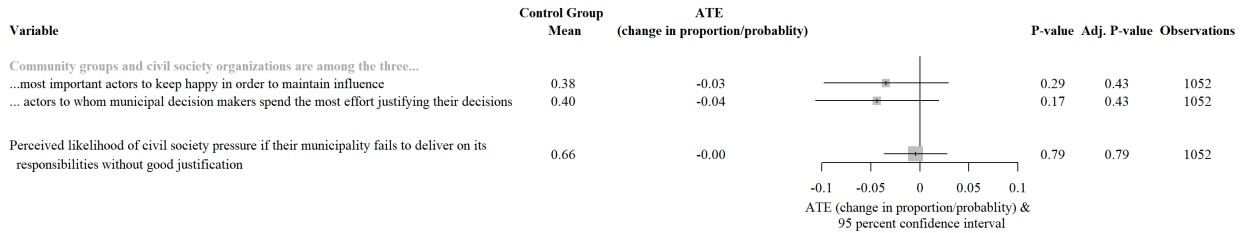
For all three variables pertaining to Hypothesis 3, we estimate average treatment effects via OLS regression, which we gives the most transparent and most easily interpretable results. All regressions include fixed effects for the randomization strata (region and anticipated security situation) and indicators for the type of decision maker (mayor, deputy mayor, opposition leader or SG). Since the dependent variable is binary for the first two outcome variables (in the order presented in Table 6) and constrained to the $[0,1]$ interval in the case of the third outcome variable, but estimated via OLS regression, we present bootstrapped standard errors and confidence intervals, adjusted for clustering at the municipality level.

Figure 17: Distribution of outcome variables for Hypothesis 3 (pooled data, treatment and control groups)



6.3.3 Unblinded results for Hypothesis 3

Figure 18: Hypothesis 3 - Results Summary



6.3.4 Alternative outcome indicators and estimation strategies considered during blind analysis

Due to the limited statistical power of our experiment and multiple testing concerns, we do not report effects on whether municipal decision-makers actually experienced civil society pressure, public protest or other forms of pressure or accountability demands, even though such data are available. We consider the anticipation of accountability pressure in potentially counterfactual situations to be of greater theoretical interest, because if the anticipation of civil society pressure is sufficient to influence municipal decision-makers' behavior, actual activism to pressure the municipal government would be off the equilibrium path. Conversely, if municipal decision-makers did experience civil society pressure, they should be more likely to also anticipate it. Therefore, we believe that the chosen indicators suffice to evaluate Hypothesis 3.

With regard to the first two indicators (whether community groups and civil society organizations rank among the top three local actors to keep happy/justify decisions to), we decided not to disaggregate further whether they rank first, second or third. The reason is that community groups/civil society rank among the top three in only just over a third of the municipalities. Therefore, we consider it sufficiently interesting whether the treatment causes additional decision-makers to rank community groups/civil society to rank among the most important actors. We also believe that a binary outcome variable will be easier to interpret for the average user or reviewer of our work, as compared to ordered outcome categories.

6.4 Hypothesis 4: 3PPP increases municipal decision makers’ awareness of performance metrics.

6.4.1 Data and methods for Hypothesis 4

To evaluate Hypothesis 4, we test if the CBO incentive scheme causes greater awareness of performance metrics among municipal decision-makers and more accurate beliefs about municipal performance. We carried out a knowledge test with the surveyed decision-makers. The knowledge tests has two parts: awareness of performance indicators and awareness of the own municipality’s performance relative to other municipalities. We separately report aggregate results for each part of the knowledge test. We aggregate the knowledge test data as follows:

- For awareness of indicators, we count the number of SUPERMUN indicators respondents were able to correctly recall from memory.
- For awareness of the municipality’s relative performance, we ask municipal decision-makers for each group of indicators (education, health, water and sanitation etc.) what quintile their municipality ranks in. This information is disclosed on the municipal performance scorecards that were disseminated in both treatment and control municipalities, but we ask respondents to give their best guess from memory, rather than looking up this information. Our outcome metric is the sum of absolute deviations between the quintile reported by the respondents and their municipality’s actual percentile, across all groups of indicators.

Figure 19: Awareness of municipal performance indicators (pooled data, treatment and control groups)

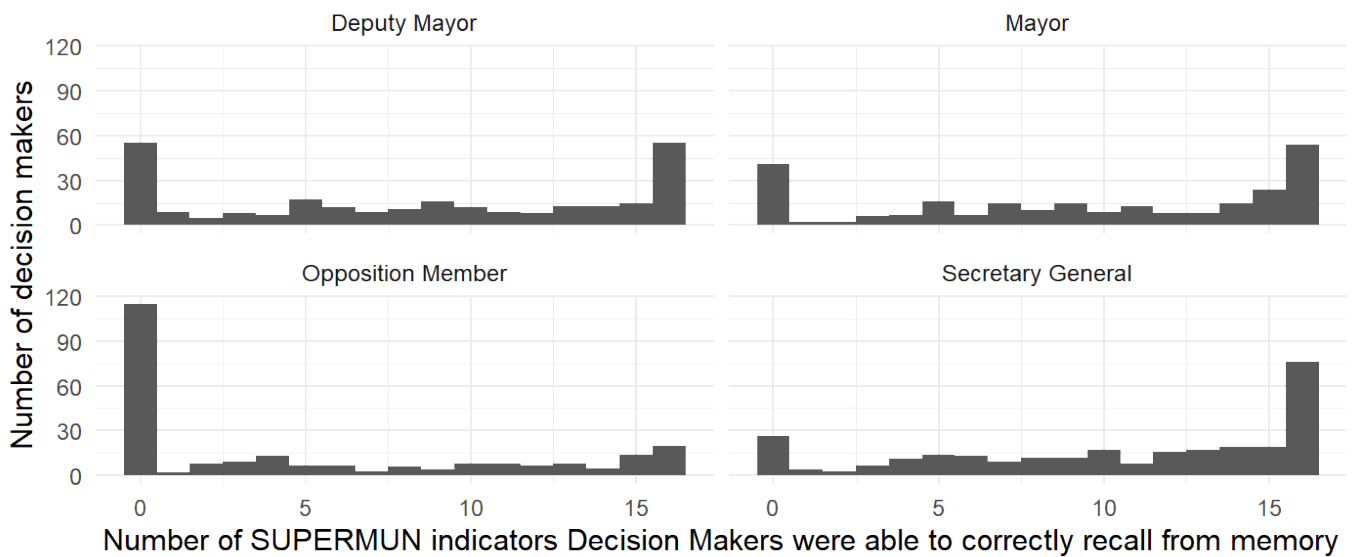
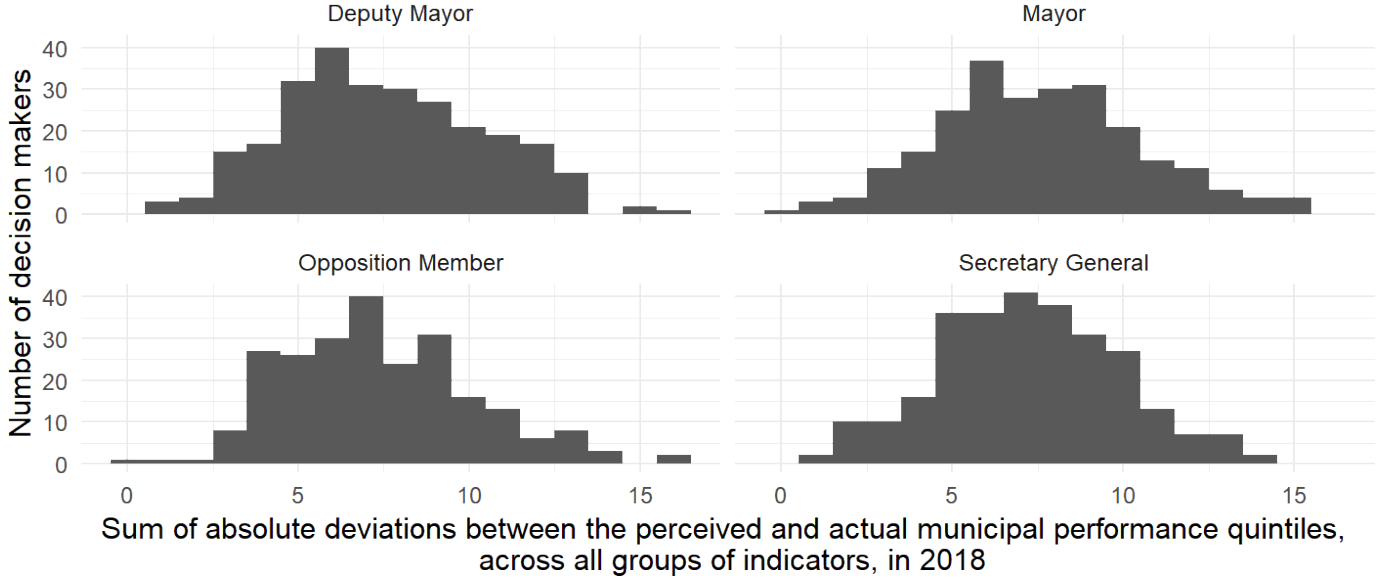


Figure 20: Accuracy of municipal decision-makers' beliefs (pooled data, treatment and control groups).



6.4.2 Estimation strategy for Hypothesis 4

We estimate effects on awareness of SUPERMUN indicators via Tobit regression, due to the high number of municipal decision-makers who are able to recall either all of the indicators, or none of them. Effects on the accuracy of municipal decision-makers' beliefs are estimated via OLS regression. All specifications include fixed effects for randomization blocks and errors are adjusted for clustering by municipality when data from different types of decision-makers are pooled. We also report results by type of decision-maker.

6.4.3 Unblinded results for Hypothesis 4

	<i>DV: Number of performance indicators remembered</i>				
	(1) All	(2) Mayors	(3) Deputy Mayors	(4) Opposition Leaders	(5) Secretary Gen- eral
<i>Control group mean</i>	7.92	8.91	7.32	4.91	10.3
Treatment	1.29* (0.715)	1.6 (1.1)	3.21** (1.26)	0.606 (1.58)	0.00561 (0.92)
Observations	1053	252	274	244	283
Censored at 0	237	41	55	115	26
Censored at 16	1053	252	274	244	283
Clusters (Municipalities)	297				

Coefficients from Tobit regressions with fixed effects for randomization blocks. Standard errors in parentheses, adjusted for clustering by municipality in column (1). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided).

Table 7: Effects on municipal decision-makers' awareness of performance indicators

*DV: Sum of absolute deviations between decision makers' perception of
and the actual municipal performance quintile*

	(1) All	(2) Mayors	(3) Deputy Mayors	(4) Opposition Leaders	(5) Secretary Gen- eral
Control group mean	7.37	7.3	7.67	7.2	7.28
Treatment	0.118 (0.214)	0.386 (0.378)	-0.167 (0.386)	0.239 (0.38)	0.00419 (0.346)
Observations	1026	244	269	237	276
Clusters (Municipalities)	297				

Average treatment effects estimated via OLS regression with fixed effects for randomization blocks. Column (1): Cluster-robust standard errors in parentheses. Columns (2)-(5): HC3 standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided).

Table 8: Effects on the accuracy of decision-makers' beliefs about municipal performance

6.4.4 Alternative outcome indicators and estimation strategies considered during blind analysis

No alternative outcome indicators needed to be considered during the blind analysis, because both indicators exhibit sufficient variation and data quality. Going forward, we are planning to consider potential covariate adjustments, e.g. for baseline performance quantile, pre-baseline exposure to SUPERMUN scorecards as part of the scorecard RCT, and individual-level predictors of test performance, such as age. Data on education levels are missing due to a questionnaire programming error.

6.5 Hypothesis 5: 3PPP increases municipal performance.

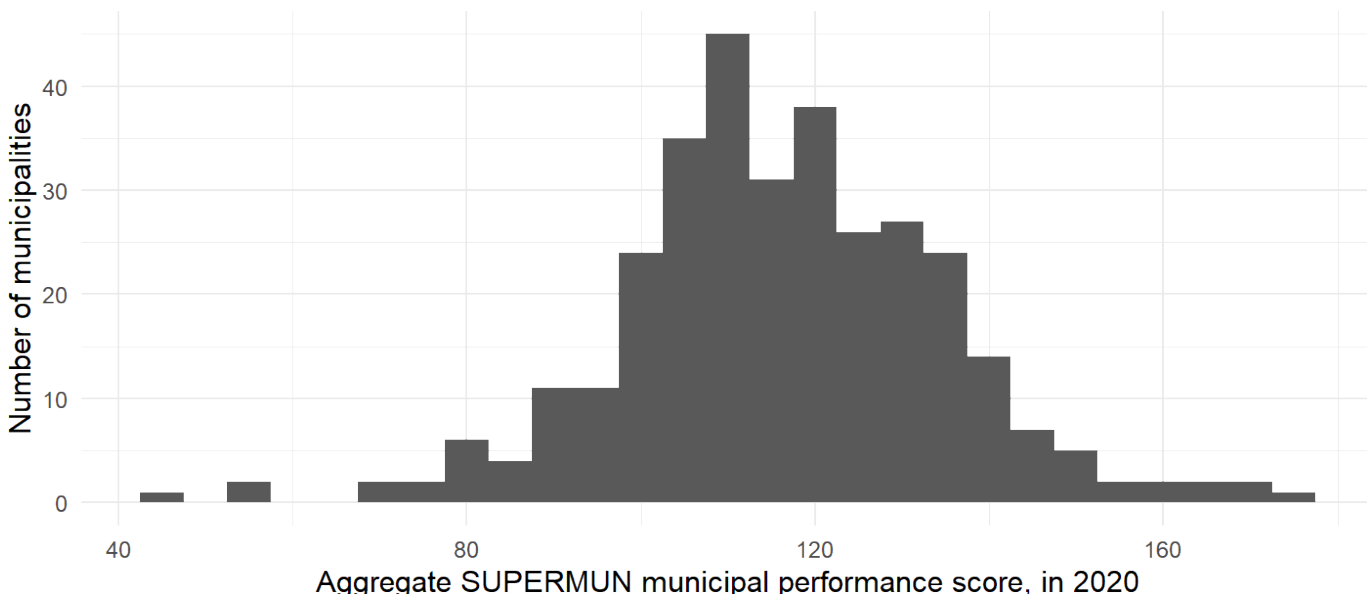
6.5.1 Outcomes of interest for Hypothesis 5

Our main outcome of interest for Hypothesis 5 is a municipality’s aggregate SUPERMUN municipal performance score in 2020. This score is a function of 16 performance indicators (nine indicators of municipal service quality and seven indicators of institutional capacity) and the main variable used to calculate CBOs’ payoff.

Our analysis focuses on the aggregate performance scores, rather than the underlying individual performance indicators, because CBOs’ strategies of influence and municipalities constraints and capacities for improvement are likely to vary. Consequently, we expect the CBO incentive scheme to affect different performance indicators in different communes in ways that are not easily predictable with the available pre-treatment data. For any individual performance indicator, average improvements across all municipalities could be small. Across indicators and municipalities, however, small improvements could add up to a sizeable average effect.

From a policy perspective, the aggregate scores are of greater relevance than the individual indicators, given that increases in effort or resources committed towards one indicator might take effort or resources away from other indicators. As a more comprehensive outcome metric capturing different dimensions of municipal performance, increases in the aggregate performance scores are more likely to reflect actual increases in effort on the part of municipal administrations than individual components of the aggregate performance scores.

Figure 21: Distribution of SUPERMUN municipal performance scores in 2020 (pooled data, treatment and control groups)



6.5.2 Estimation strategy for Hypothesis 5

We estimate standardized average treatment effects on municipalities’ performance scores via linear regressions of the aggregate SUPERMUN municipal performance scores in 2020 on the treatment indicators. We include the 2018 baseline SUPERMUN performance scores, as well as baseline conflict exposure (measured by ACLED data on total conflict fatalities in 2018) as pre-treatment covariates, along with fixed effects for randomization blocks.

The two pre-treatment covariates are standardized within treatment and control groups (to have zero mean and a standard deviation of one) and interacted with the treatment indicator, so that the main effect on the treatment indicator is an unbiased estimate of the average treatment effect. The dependent variable is standardized by subtracting the control group sample mean and dividing by the control group sample standard deviation.

In the absence of any empirical indication that the treatment could have had negative impacts on municipal performance (based on municipal decision-makers' evaluations of this possibility in the survey), we conceive of Hypothesis 5 as a one-sided alternative to the null hypothesis of no effect.

Missing values

Due to the security situation, not all SUPERMUN indicators could be collected in all municipalities in 2018 and 2020. However, if individual components are missing, the overall performance score can also not be calculated, which limits the number of observations available to test Hypothesis 5 to 258 municipalities. To address this issue, we impute missing baseline and endline values of individual indicators by using the prior-year values (i.e. 2017 and 2019 values, respectively). Imputed missing indicators are used together with the non-missing other indicators to calculate the overall performance scores. We separately report estimation results without any adjustment, with imputed baseline values, and with imputed baseline and endline values.

Heterogeneous treatment effects

We do not have strong expectations regarding treatment effect heterogeneity along the pre-treatment conflict dimension. One reason is that conflict patterns changed between 2018 and 2020, another reason is that the most conflict-affected areas are already excluded from the experiment. In conflict-affected areas, municipalities face greater constraints in fixing performance shortfalls, but that does not necessarily imply that accountability pressure from CBOs could not contribute to preventing further deterioration.

Our prior expectations regarding treatment effect heterogeneity between high- and low-performing municipalities. On the one hand, municipal governments that face tighter external constraints on their performance are likely to exhibit lower performance at baseline, while simultaneously having less potential for performance improvements. This might diminish potential treatment effects.

If we find significant heterogeneity in treatment effects on CBO involvement (Hypothesis 1) between rural and urban municipalities, we plan to also test for heterogeneous effects of urban/rural status on municipal performance outcomes.

6.5.3 Unblinded results for Hypothesis 5

Table 9 reports average treatment effect estimates on aggregate municipal performance scores, on which the CBOs' payoff, and our test of Hypothesis 5 is based. In Figure 22, we additionally report average treatment effects on the individual components of the SUPERMUN performance scores. This analysis should be considered exploratory, as we do not have specific expectations regarding particular performance indicators the CBOs might focus on. Rather, we expect that CBOs in different municipalities may focus on different aspects of municipal performance, so that average effects on individual indicators might not be detectable, even if there is an effect on overall performance scores.

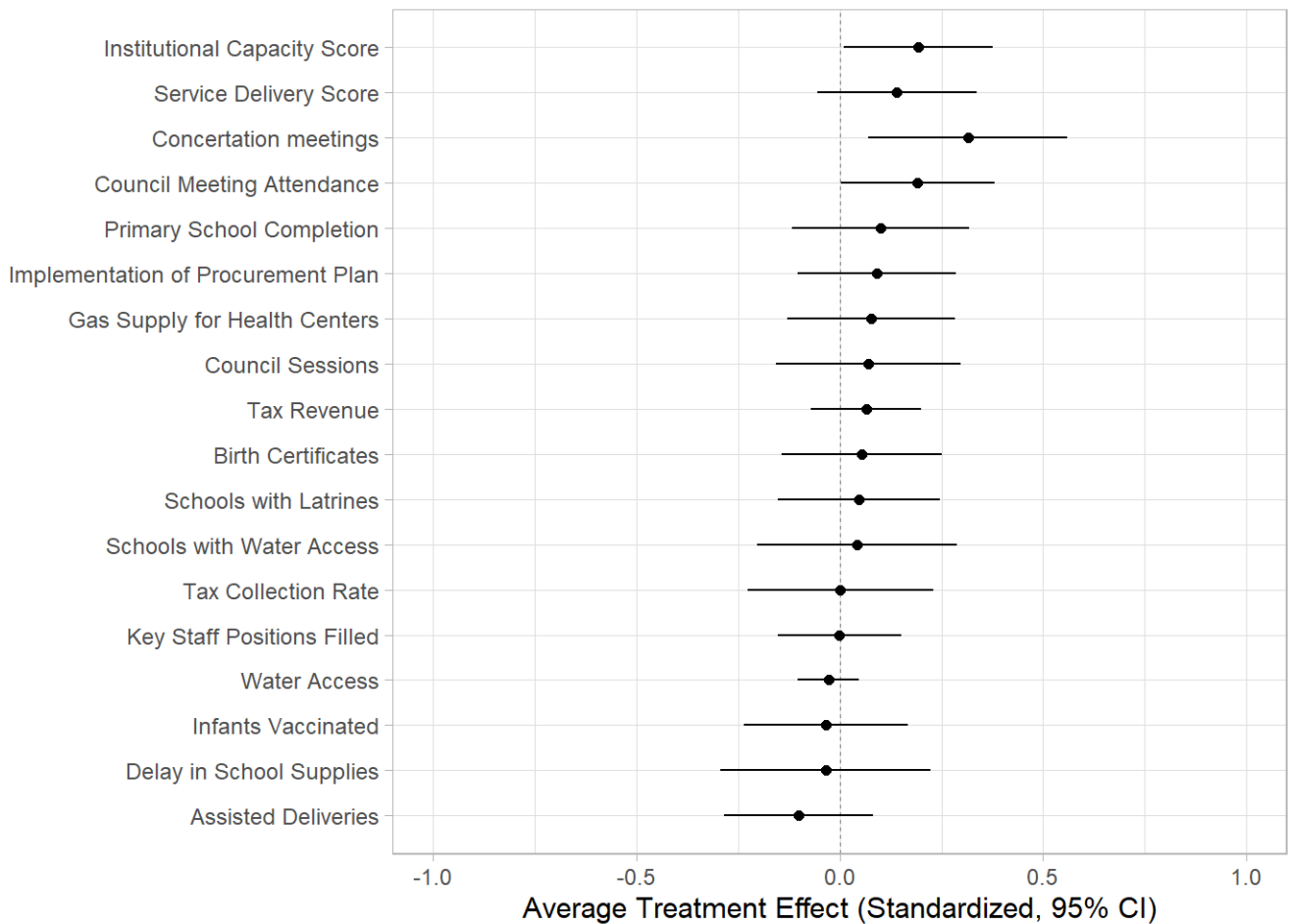
DV: SUPERMUN municipal performance score (standardized)

	(1) No adjustment	(2) Missing baseline values imputed	(3) Missing baseline and endline values imputed
Standardized ATE	0.22** (0.095)	0.24** (0.095)	0.22** (0.094)
Baseline performance	0.5*** (0.082)	0.49*** (0.079)	0.49*** (0.079)
Treatment × Baseline performance	-0.12 (0.11)	-0.11 (0.11)	-0.13 (0.11)
Baseline conflict exposure	-0.029 (0.035)	-0.024 (0.034)	-0.016 (0.034)
Treatment × Baseline conflict exposure	0.12* (0.067)	0.11 (0.067)	0.1 (0.065)
Observations (Municipalities)	280	287	292

All effects are estimated via OLS regressions with fixed effects for randomization blocks. In columns (2) and (3), missing values on individual performance indicators have been replaced with prior-year values and used together with the other, available indicators to calculate the aggregate performance score. HC2 standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided).

Table 9: Effects on SUPERMUN municipal performance scores

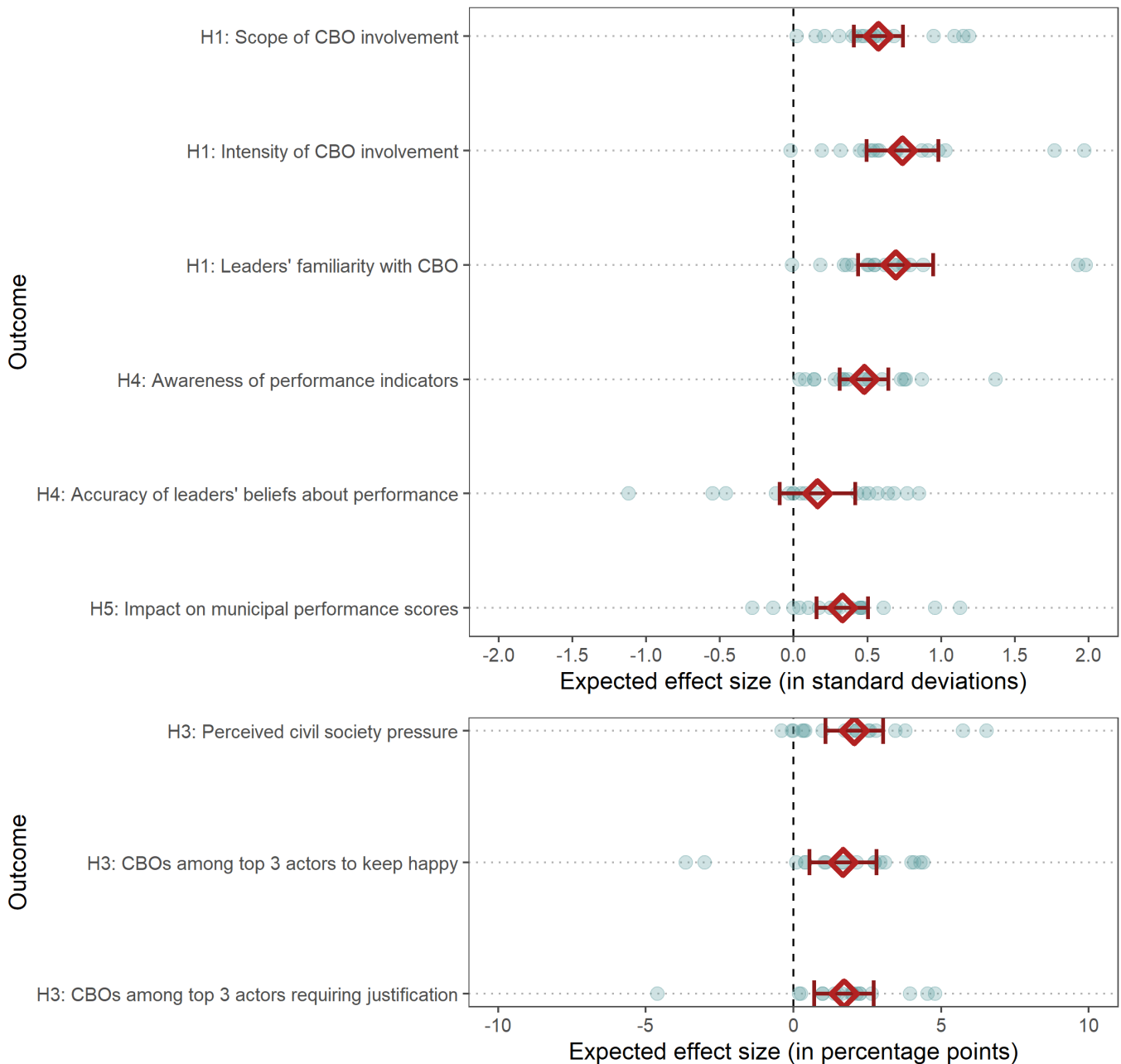
Figure 22: Estimated Treatment Effects on Individual Performance Indicators



7 Expectations Survey

After presenting the results-blind report in research seminars to audiences with pertinent research backgrounds, we conducted a survey of seminar participants, asking them for their expectations regarding the effect sizes of the 3PPP treatment on each of the outcomes considered in the blind analysis.¹ Regarding Hypotheses 1, 3, 4 and 5, most respondents expect positive and relatively large treatment effects, for all outcomes except the accuracy of decision makers' beliefs (Figure 23). Additionally, two thirds of respondents expect that the experiment will be able to reject that there are any adverse effects on the internal functioning of CBOs (Hypothesis 2).

Figure 23: Researcher expectations regarding effect sizes



¹The survey was mildly incentivized: The respondent whose estimates were closest on average to the true effect would win a crate of banana beer or a 20 EUR gift card, if they optionally left a contact e-mail. The prize notwithstanding, respondents were guaranteed that their answers would only be published in anonymized format.

8 Conclusions

Through a large-scale randomized controlled trial in Burkina Faso, we test if third-party performance pay (3PPP) can increase the involvement of community-based organizations (CBOs) in municipal governance, improve municipal government accountability, and ultimately increase municipal government performance. Under the 3PPP intervention, selected community-based organizations (CBOs) were promised a variable cash grant that was proportional to changes in their municipal government’s performance scores over a two-year period. Partner CBOs could decide on their own if they wanted to take any initiative towards increasing their municipal government’s performance and how they would do so. They would remain eligible for the cash grant, even if they did nothing at all.

We conducted a results-blind analysis to minimize experimenter bias and improve research transparency. The results-blind analysis enabled us select hypothesis tests, data processing and estimation strategies with knowledge of the data, but without knowledge of the actual results. All estimations and hypothesis tests presented in this report, as well as the preceding data cleaning and data processing steps, were chosen in a results-blind manner and are therefore neither consciously, nor unconsciously motivated by how they might affect the outcome of hypothesis tests or the readers’ interpretation of the results.

Upon unblinding of the study, we find that third-party performance pay (3PPP) was effective at stimulating the voluntary involvement of community-based organizations in municipal governance and increased overall municipal performance scores by an estimated 0.22 standard deviations.

- CBOs in the treatment group significantly increased the scope of their participation in municipal governance and additionally interacted more frequently with local service providers. Municipal decision-makers report greater familiarity and interaction with the treatment CBOs, compared to control CBOs.
- We find no evidence of adverse effects on the internal functioning of treatment CBOs. Specifically, no differences were observed with regard to collective action capacity, internal cohesion, perceived organizational capture and indicators of organizational capacity, except for a positive effect on the organizations’ COVID-19 adaptability.
- Contrary to expectations, we also find no evidence that 3PPP increased municipal decision-makers’s perception of accountability pressures.
- We find limited evidence of increases in municipal decision-makers’ awareness of performance indicators, largely driven by changes in deputy mayors’ awareness, but no detectable differences in the accuracy of decision-makers’ performance self-assessments. This was to be expected by design, as municipal decision-makers’ access to performance information was held constant across treatment and control municipalities.

The results suggests that 3PPP may have incentivized constructive forms of involvement, such as gentle nudging and collaborative problem-solving between CBOs and municipal authorities, more than confrontational demands for accountability that would have increased decision-makers’ perception of accountability pressures. However, a possibility remains that the decision-makers’ perceptions of accountability pressures were mismeasured or operationalized too narrowly.

In sum, the experiment demonstrates that 3PPP can be an effective and scalable way of incentivizing the involvement of grassroots civil society in local governance and improving local government performance. These results are particularly notable in two respects: First, the experiment took place under very adverse conditions, characterized by high levels political instability, the outbreak of the COVID-19 pandemic, and escalating armed conflict. Second, the financial incentives to CBOs were small by design and further diminished by exogenous declines in municipal performance. Yet,

we find that even such limited incentives were effective at increasing partner CBOs' involved in municipal governance and improving overall municipal performance. This points to a potentially high cost effectiveness of 3PPP.

9 References

- Andrabi, T., Das, J., Khwaja, A. I. (2017). Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets. *American Economic Review*, 107(6), 1535-1563.
- Ariely, D., Bracha, A., Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review*, 99(1), 544-55.
- Asadi-Aliabadi, M., Karimi, S. M., Tehrani-Banihashemi, A., Mirbaha-Hashemi, F., Janani, L., Babae, E., Nojomi, M., Moradi-Lakeh, M. (2022). Effectiveness of pay for performance to non-physician health care providers: A systematic review. *Health policy*. <https://doi.org/10.1016/j.healthpol.2022.03.007>.
- Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R., Khemani, S. (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy*, 2(1), 1-30.
- Banerjee, A. V., Duflo, E., Glennerster, R. (2008). Putting a band-aid on a corpse: incentives for nurses in the Indian public health care system. *Journal of the European Economic Association*, 6(2-3), 487-500.
- Barr, A., Packard, T. G., Serra, D. (2012). Participatory accountability and collective action: evidence from field experiments in Albanian schools.
- Belle, N., Cantarelli, P. (2015). Monetary incentives, motivation, and job effort in the public sector: An experimental study with Italian government executives. *Review of Public Personnel Administration*, 35(2), 99-123.
- Björkman, M., Svensson, J. (2009). Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda. *The Quarterly Journal of Economics*, 124(2), 735-769.
- Christensen, D.; Dube, O., Haushofer, J., Siddiqi, B., Voors, M. (2021). Building Resilient Health Systems: Experimental Evidence from Sierra Leone and The 2014 Ebola Outbreak, *The Quarterly Journal of Economics*, 136(2), 1145–1198.
- Deci, E. L., Koestner, R., Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, 125(6), 627.
- Diaconu K., Falconer J., Verbel A., Fretheim A., Witter S. (2021). Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database of Systematic Reviews*. 2021(5). Art. No.: CD007899. DOI: 10.1002/14651858.CD007899.pub3.
- Ensminger, J. (2012). Inside Corruption Networks: Community Driven Development in the Village. Unpublished manuscript.
- Fang, M., Gerhart, B. (2012). Does pay for performance diminish intrinsic interest? *The International Journal of Human Resource Management*, 23(6), 1176-1196.
- Fiala, N., Premand, P. (2018). Social accountability and service delivery: Evidence from a large-scale experiment in Uganda. World Bank Policy Research Working Paper No. 8449, Washington, DC: World Bank.
- Frey, B. (2017). Policy consequences of pay-for-performance and crowding-out. *Journal of Behavioral Economics for Policy*, 1(1), 55-59.
- Fuster, A., Meier, S. (2010). Another hidden cost of incentives: The detrimental effect on norm enforcement. *Management Science*, 56(1), 57-70.

- Humphreys, M., Sanchez de la Sierra, R., van der Windt, P. (2013). Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration. *Political Analysis*, 21(1), 1-20. doi:10.1093/pan/mps021
- Khwaja, A. I. (2004). Is increasing community participation always a good thing? *Journal of the European Economic Association*, 2(2-3), 427-436.
- Klein, J. R., Roodman, A. (2005). Blind Analysis in Nuclear and Particle Physics. *Annu. Rev. Nucl. Part. Sci.*, 55, 141–163. doi: 10.1146/annurev.nucl.55.090704.151521
- Lierl, M., Holmlund, M.. (2019). Performance Information and Voting Behavior in Burkina Faso’s Municipal Elections: Separating the Effects of Information Content and Information Delivery. In Dunning, T., Grossman, G., Humphreys, M., Hyde, S., McIntosh, C. (Eds.), *In Information, Accountability, and Cumulative Learning : Lessons from Metaketa I* Cambridge University Press.
- MacCoun, R., Perlmutter, S. (2015). Blind analysis: Hide results to seek the truth. *Nature*, 526, 187–189. <https://doi.org/10.1038/526187a>
- Mansuri, G., Rao, V. (2013). Can participation be induced? Some evidence from developing countries. *Critical Review of International Social and Political Philosophy*, 16(2), 284-304.
- Mansuri, G., Rao, V. (2013). *Localizing Development: Does Participation Work?*
- Nyquist, M., De Walque, D., Svensson, J. (2017). Experimental evidence on the long-run impact of community-based monitoring. *American Economic Journal: Applied Economics*, 9(1), 22-69.
- Olken, B. A. (2007). Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy*, 115(2), 200-249.
- Paul, E., Bodson, O., Ridde, V. (2021). What theories underpin performance-based financing? A scoping review. *Journal of Health Organization and Management*, 35(3), 344-381.
- Paul E., Albert L., Bisala BN’S., et al. (2018). Performance-based financing in low-income and middle-income countries: isn’t it time for a rethink? *BMJ Global Health*, 3:e000664. DOI:10.1136/bmjgh-2017-000664
- Platteau, J. P., Gaspart, F. (2003). The risk of resource misappropriation in community-driven development. *World development*, 31(10), 1687-1703.
- Raffler, P., Posner, D. N., Parkerson, D. (2022). *Can Citizen Pressure Be Induced to Improve Public Service Provision?*. Working Paper, Harvard University.
- Singh, N. S., Kovacs, R. J., Cassidy, R., Kristensen, S. R., Borghi, J., Brown, G. W. (2021). A realist review to assess for whom, under what conditions and how pay for performance programmes work in low- and middle-income countries. *Social Science and Medicine*, 270(2), 113624. <https://doi.org/10.1016/j.socscimed.2020.113624>.