Community based monitoring and public service delivery - Impact, and the role of information, deliberation, and jurisdictional tier

Nassul Kabunga*, Caroline Miehe†, Tewodaj Mogues‡, Bjorn Van Campenhout§†

June 12, 2020

Abstract

To improve public service delivery, the Government of Uganda organizes community forums—popularly known as barazas—where citizens receive information from government officials, and get the opportunity to directly engage with them. We run a cluster randomized control trial to assess the impact of this policy intervention on public service delivery in agriculture, health, education, and infrastructure. Using a factorial design, we further test the relative importance of the two main components of the intervention—information provision and citizen engagement. we also compare the effectiveness of barazas organized at the district level to the effectiveness of barazas organized at the sub-county level. Using a strictly pre-registered confirmatory analysis, we find no impact of the intervention on general public service delivery, but there are some indications that sub-county level barazas increase outcomes in the agricultural sector. A more exploratory part that looks at individual outcomes, potential mechanisms, and heterogeneous treatment effects suggests a range of localized impacts of barazas.

1 Introduction

In 2015, we designed a study aimed at evaluating the effectiveness of community based monitoring and accountability fora, also known as barazas, in Uganda. The baraza programme, an initiative of the president of Uganda and implemented by the Office of the Prime Minister (OPM), was designed to improve

*Uganda Bureau of Statistics, Kampala, Uganda

†LICOS Centre for Institutions and Economic Performance, KULeuven, Leuven, Belgium

‡International Monetary Fund, Expenditure Policy Division, Fiscal Affairs Department, Washington DC, USA

§Development Strategy and Governance Division, International Food Policy Research Institute, Leuven, Belgium - corresponding author: b.vancampenhout@cgiar.org

public service delivery by enhancing public involvement in holding the government accountable for service delivery in relation to resources spent (OPM, 2013). The study had several objectives. First, it wanted to establish, in a rigorous way, if the program had an impact on public service delivery and associated outcomes. A second objective of the study was to compare the effectiveness of barazas organized at lower administrative levels (the sub-county) to that of barazas that are organized at a more aggregate level (the district). Third, the project also set out to assess the relative importance of the two main components of a generic baraza event: using a two-by-two factorial design, it differentiates between the impact of (1) providing citizens with information related to budgeting and spending and planning, and (2) the impact of allowing citizens to engage with public servants and politicians in moderated questions-and-answers sessions. Towards the end of 2015, a baseline survey involving more than 12,500 households and 400 government officials was conducted. After completion of the baseline, we trained local government officials and community leaders to ensure adherence to the intervention protocols, and the interventions were rolled out by the OPM.

While the study was initially assumed to take about 2 years, OPM faced various challenges that affected the timely roll-out of the intervention, including budgetary constraints and disruptions related to the general elections of 2016. Four years after the baseline survey, with about 50 percent of the planned interventions completed, a trade-off needed to be made between waiting for the remaining barazas to be completed or conducting the end-line after partial roll-out. It was decided to proceed with end-line data collection and employ various strategies to diagnose and control for potential selection bias that may have been introduced by the partial roll-out. Early 2020, endline data was collected on 6,700 households and 260 government officials.

A confirmatory analysis, strictly adhering to a pre-registered pre-analysis plan that summarizes a range of outcomes into four different families of outcomes corresponding to the main sectors—agriculture, infrastructure, health and education—and one overall index shows little evidence that the baraza intervention had an impact on public service delivery or its associated outcomes. The only exception is service delivery in the agricultural sector, where we find that sub-county level barazas have a positive impact, and that this impact is clearly superior to the (lack of) impact associated with district level barazas.

In a more exploratory analysis, we first look at individual outcomes in each of the four sectors in more detail. We then look for changes in behavior that is explicitly targeted by the intervention, such as information seeking behavior and contributions to the common good. Finally, we provide results on changes in the perception of citizens on a range of problems that came up in discussions with stakeholders during qualitative fieldwork. We also explore heterogeneity in the treatment effects arising from differences in: timing of the interventions; political connections of the sub-counties; baseline levels of elite capture; ethnic fractionalization; and distance of the household to the sub-county headquarters.

The exploratory analysis adds nuance to the conclusion arrived at in the confirmatory part. The sector level aggregate outcomes hide significant impacts

on access to drinking water and schooling infrastructure. But there are also signs that the aggregate treatment effects mask significant heterogeneity. For instance, the impact of sub-county barazas on access to extension seems to be driven particularly by households that live close to the sub-county headquarters. The deliberative component of the baraza seems to increase service delivery in the area of infrastructure and reduces distance to the nearest public health facility, particularly if the sub-county has political connections.

There have been several studies that look at the impact of community involvement on public service delivery, many of them using Uganda as a case. A landmark study is Björkman and Svensson (2009), who analyse the impact of a community driven local accountability project in primary health care provision in Uganda. They find that the intervention resulted in significant improvements in health care delivery, utilization, and health outcomes (most notably child mortality and weight-for-age z-scores) after one year, and confirm in Björkman Nyqvist, de Walque, and Svensson (2017) that these effects also hold in the longer run. More recently, however, Raffler, Posner, and Parkerson (2018) come to more nuanced conclusions when testing an intervention closely modeled on the one of Björkman and Svensson (2009) and in a similar setting. The study, involving a three wave panel of more than 14,000 households and a factorial design to break down the intervention into its two most important components similar to what we use, shows that the intervention was able to change the behavior of public service providers. However, it does not find that the intervention directly increased community monitoring nor generated improvements in health outcomes in the short run.

Our study contributes to this literature in various ways. First, this study is one of few that considers the role of administrative placement on the effectiveness of community monitoring. The level at which the intervention occurs may affect its effectiveness in opposing ways. Interventions at a more local level may result in more relevant issues being scrutinized. However, qualitative explorations suggest that often, issues raised in lower level barazas fall under the responsibility of higher levels of government or other institutions that are beyond the operational jurisdiction of the officials that are present at the baraza (Van Campenhout et al., 2018). This may be less of a problem when barazas are organized at district level. Most other studies consider interventions that are placed at fairly local levels. For instance, the intervention in Raffler, Posner, and Parkerson (2018) is implemented in health centers and their associated catchment areas, defined as the three villages closest to the health center.

Second, our study evaluates the impact of a government initiative, which may instigate an entirely different set of dynamics than interventions that are organized by local bodies or international NGO's. For instance, many of the actors involved may find that NGOs are not mandated when it comes to public services such as health and education. Furthermore, it is likely to be easier to re-allocate resources to problems identified during barazas when they are organized by the government. This is also consistent with suggestive evidence in Raffler, Posner, and Parkerson (2018), who find that the presence of sub-county officials during their community based monitoring intervention boosted

the impact of the intervention. However, effects may also work in the opposite direction. For example, an intervention to reduce absenteeism in government public health facilities in India was initially very successful, but ceased to have any impact after the local bureaucracy started providing official excuses for most of the nurses' absences (Banerjee, Duflo, and Glennerster, 2008). Most of the other studies that are closest to our study use NGOs as implementing partners (eg. Björkman and Svensson, 2009; Raffler, Posner, and Parkerson, 2018).

Third, baraza's take a comprehensive, multi-sector approach, enabling cross-sectoral planning and potentially allowing for re-allocation of resources across sectors. Some of the problems most mentioned by users, such hygiene in health centers or accessibility, involve cooperation between heads of different sectors (eg. health and infrastructure to get access to water in health centers). Bringing sector heads together and confronting them with the priorities of citizens may increase information sharing and cooperation between them. Most other studies focus on a single sector. Health in particular seems to be a popular sector for community monitoring interventions (eg. Arkedis et al., 2019; Björkman and Svensson, 2009; Raffler, Posner, and Parkerson, 2018)

Finally, we evaluate a high-profile policy intervention that receives broad support within government and civil society, and among opinion leaders and citizens in Uganda. Evaluating policy interventions has it challenges, and this one is no exception. As a result, such research has become rare and present day rigorous impact evaluations often bypass the political resistance to randomization among governments, implementing agencies, and beneficiaries. As a result, the nature of the partners has changed (NGOs rather than governments) and interventions have become "relatively trivial" (de Souza Leão and Eyal, 2019).

The paper starts with providing some background on the baraza intervention, followed by the research questions. We then present the experimental design and describe the treatments. Next is a section on the implementation challenges that prompted some changes to the original design and updated power calculations. The following section presents results, starting with an aggregate analysis that closely adheres to a pre-registered report. This is followed by an in-depth analysis of individual outcomes. We also explore some of the assumed impact pathways, and investigate treatment heterogeneity in a separate section. The last section concludes.

2 Background of the baraza impact evaluation

In the mid-1980s, after attaining relative stability, the Government of Uganda, supported by development partners, initiated far reaching liberalization efforts and introduced a decentralized system of governance (Francis and James, 2003). Decentralization was considered a suitable mechanism for improving efficiency, effectiveness and inclusiveness of public service delivery. It was assumed it would lead to the formulation of services that are more aligned to citizens' needs and bring representative governance closer to citizens. However, in order to work, decentralization requires a certain level of citizen empowerment, with citizens

actively participating in planning, implementation, monitoring and evaluation of development interventions in their locations, so as to improve accountability and responsiveness of local leaders and service providers.

Until recently, this empowerment component was lacking in Uganda's decentralization process. As a result, the realization of benefits of decentralization in Uganda has been greatly affected by ineffective monitoring and weak accountability mechanisms, especially with respect to beneficiaries holding the service providers accountable (Björkman and Svensson, 2009; Reinikka and Svensson, 2004). In response to this, the Government of Uganda, under the stewardship of the Office of the Prime Minister (OPM), initiated community based monitoring and accountability fora (or citizen barazas) in 2009 with the general objective of "enhancing public involvement in holding the government accountable for service delivery in relation to the resources spent" (OPM, 2013).

Barazas were conceived as platforms for enhancing information sharing between policy makers (the client), public servants (the implementer), and beneficiaries of public goods and services (the users). In addition, barazas were designed to provide the opportunity to users to ask questions to the policy makers and public service providers, and deliberate among themselves. It is expected that, ultimately, barazas will contribute to effective monitoring, and increase accountability and transparency among all stakeholders.

Barazas have been implemented in Uganda for about 10 years. They were first piloted in the financial year 2009/10. Since then, efforts have been underway to roll out barazas in all sub-counties in the country. During the full-scale implementation phase in the financial year 2010/2011, 16 more sub-counties in 8 districts had held a baraza event. And, by the last quarter of 2011/2012, 267 out of the country's total of 1,340 sub-counties, spread over 112 districts, had held a baraza meeting. At the beginning of the 2012/2013 financial year, however, changes in implementation were suggested: subsequent barazas would target districts instead of sub-counties to increase participation at a higher level and reduce costs.

A typical baraza is initiated from the center, with the OPM mobilizing district and sub-county officials. These include the Chief Administrative Officer (CAO) as the head of public service providers at the district level, the Resident District Commissioner (RDC) as a direct representative of the president, the District Local Council Chairperson (LC5) as the representative of political leadership at the district level, and the various sector heads responsible for public services in each sector (agriculture, education, infrastructure and health). Especially for barazas organized at the sub-county level, the sub-county level equivalents of the CAO (the sub-county chief) and the LC5 (the sub-county chairperson, or LC3) also have important roles. OPM, in consultation with the district leaders (RDC, CAO and the LC5) and other stakeholders, agree on the date and a neutral location. Again, in consultation with the district leaders, a viable moderator and an interpreter into the local language (where applicable) are identified to guide the baraza forum. Village mobilizers and community resource persons are used to publicize the event. These community mobilization efforts are further reinforced by adverts in the local media in the form of radio

announcements, printed banners, posters and fliers, and mobile public address systems, a few days before the baraza event.

A baraza meeting is chaired by the Office of the RDC in each district. In front of the audience, including local citizens, invited opinion leaders, elders, and journalists, the RDC seeks feedback from each head of major sectors. Sector heads are required to present what services were planned to be delivered in the sub-county, what was actually delivered and in what quantity and quality, and what issues and challenges have emerged and what is the way forward. The RDC then seeks feedback from citizens on whether what has been presented is what they understand was planned for and actually implemented in different locations. Sector heads are then given another opportunity to clarify on or react to any issues raised by the citizens. At the end of the process, the RDC makes a report to the OPM, indicating issues that arose in the baraza meeting. This report particularly points out policy and program implementation weaknesses and challenges, which is then expected to further feed into the general government performance management system. Usually, a minister of state is also present at the baraza. Baraza events often become very emotional and the events attract considerable media attention, also from the national newspapers.

3 Research questions: impact, information versus deliberation component, and administrative placement

3.1 The impact of (sub-county level) barazas

The baraza intervention fundamentally seeks to improve public services through improving accountability of local public decision makers and service providers. The baraza intervention as conceived by the OPM is a fairly standard community based monitoring intervention that combines the provision of information with the possibility of citizens to engage with each other and with decision makers at a fairly local level. Such community based monitoring has become a popular tool to increase service delivery. However, not all such interventions appear to be successful (Olken, 2007). A first question is therefore simply related to the impact of a typical baraza intervention as originally conceived and organized by OPM. In particular, we will test the impact of sub-county level barazas that combine information provision with a deliberation component on public service delivery and associated outcomes.

3.2 The Information Mechanism

In situation characterized by incomplete and asymmetric information, targeted efforts to fill knowledge gaps can make a big difference. Indeed, the relationship between citizens and elected officials is a classic example of the principal–agent problem. Hence, providing citizens with information about the performance of

the agent is assumed to be an effective way to increase the quality of public service delivery by allowing citizens to monitor and apply pressure on underperforming politicians and civil servants (Raffler, Posner, and Parkerson, 2018). A second central question in this study therefore relates to the relative importance of the information provision component within the broader baraza intervention.

There is some evidence that channeling of information to citizens about the quantity, modality, and quality of public services, as well as about the investments and policy decisions made by politicians, bureaucrats, and service providers can increase the ability of the users to hold the leaders accountable and improve service provision. For example, Pandey, Goyal, and Sundararaman (2009) establish, using a field experiment in India, that community information campaigns about states' school management obligations had a positive impact on school performance. Grossman and Michelitch (2018) disseminate information about job performance for randomly selected Ugandan politicians. While this increases job performance for the politicians on a range of criteria, they find no impact on public service provision. Banerjee et al. (2018) find that targeted information reduces elite capture in Indonesia: mailing cards with program information to targeted beneficiaries increases the subsidy they receive from a subsidized rice program. A recent review of 48 empirical studies on the impacts of information on governance and service delivery also suggests that the availability of information alone may not suffice. Information must be deemed relevant to its recipient, and individuals must have both the power and incentives to act on the information (Kosec and Wantchekon, 2020).

The information component may also be important in managing expectations of the client. Citizens may have exaggerated beliefs about the resources at the disposal of decision makers and service providers, or they may not fully appreciate the challenges they face when doing their job. For instance, during focus group discussions, service providers mentioned that citizens sometime blame officials for things they have no control over. More in general, information may help sensitize citizens about the role of public service provision (for example, making sure boreholes are present) but that there are also limits to what citizens can expect (for instance, citizens are still required to boil water). Where information can rectify expectations, we may not find changes in the quantity or quality of public services, but we may still find changes in citizens' perceptions about the quality of these services.

3.3 The Deliberation Mechanism

There are various ways in which deliberation may increase the quality of public service delivery. First, it has a legitimizing effect on decisions arrived at in this fashion. Effective deliberation assumes equal voice of the arguments of both marginal and advantaged agents, and the role of evidence that support the positions articulated. Second, deliberation can more effectively distill social choice than simple voting and majoritarian rule, in part by building consensus both among citizens and between public servants and citizens. Third, deliberation

has been found to positively impact the vigor and breadth of subsequent citizen involvement in community affairs (Björkman Nyqvist, de Walque, and Svensson, 2017).

Deliberation may also affect information flows. In a baraza, the information component is primarily designed to inform citizens about the activities of the service providers. To some extent, citizens are passive recipients of this information, and officials report what they consider relevant, or may even attempt to misrepresent the facts. If citizens are able to engage with policy makers and civil servants, they may request particular information that is relevant to them. It may also result in information flows in the opposite direction as government officials learn about priorities and concerns of the citizens. The third key question is thus about the relative importance of the deliberation aspect of a typical baraza.

Impacts of deliberative processes have also been the subject of empirical analysis. For example, in addition to increasing community participation mentioned above, experimental evidence also shows that deliberative processes make decision outcomes less sensitive to the institution (e.g. voting) rules that bring them about (Goeree and Yariv, 2011) or may reduce the prevalence of clientelism (Fujiwara and Wantchekon, 2013).

3.4 Administrative placement

The baraza intervention can also be distinguished by the administrative level at which it is implemented. Barazas were originally implemented at the sub-county level but from 2012 onward, more and more barazas were implemented at the district level. This administrative placement dimension immediately points to a potential trade-off between attempting to achieve breadth of coverage (through district-level barazas), and depth of coverage (through sub-county-level barazas). While conducting a district-level baraza may be cheaper than conducting sub-county-level barazas in all sub-counties of that district, it is not clear a-priori how these cost savings justify potential reduction in effectiveness of district-level barazas in any given sub-county of the concerned district. Therefore, another vital aspect of this study is to assess the impact of district level barazas, and compare this to the effectiveness of barazas organized that the lower sub-county level.

Which is more effective, placement at a higher or lower level, will likely depend on the outcome and the situation. For instance, it has been argued that engaging small groups can be more effective because they can be coordinated more easily, but large groups may make more sense if the desired outcome would be enjoyed by a broader group (Donato and Mosqueira, 2016). Furthermore, action may be more likely if an issue is brought by a large group instead of a small group of people complaining about a highly localized issue (Banerjee, Deaton, and Duflo, 2004). It may also be that issues highlighted at a local level fall under the responsibility of higher level authorities and vice versa.

4 Experimental design and balance at baseline

To answer the above questions, we designed a social experiment covering districts, sub-counties, and households across the four regional blocks (Northern, Western, Central and Eastern) of Uganda. Each regional block has somewhat unique characteristics in terms of ethnicity, geographical and agro-ecological conditions, as well as cultural history. As noted in Section 2, a small share of all sub-counties, albeit located throughout all of Uganda's 112 districts across the four regions, had already received a sub-county level baraza intervention prior to the study. We thus selected our sample of districts from among 'eligible districts', and our sample of sub-counties from 'eligible sub-counties'. An 'eligible district' was defined as a district in which a district level baraza was not already implemented prior to the start of the study. An 'eligible sub-county' was defined as a sub-county to which two conditions applied: (i) a sub-county level baraza had not yet taken place, and (ii) the sub-county was in an 'eligible district'. Preliminary analysis of the baraza implementation data at the time of the start of the study indicated that there were 20 or more eligible districts per region, amounting to a total of 94 eligible districts. In each region, there were at least 147 sub-counties that had never been treated and were in eligible districts; the total of such eligible sub-counties was about 720[1].

This study employed a nested, or two-step, randomization design, illustrated in Figure 1. In a first step, we randomly allocate eligible districts to treatment and control conditions. In particular, some of the eligible districts start receiving district level barazas $\left(D^{ID}\right)$, while other districts do not receive a baraza at this level $\left(D^{0}\right)$[2]. In a second step, we proceed with all eligible sub-counties and randomly allocate each sub-county to one of four conditions in a 2 by 2 factorial design. In particular, about one quarter of all eligible sub-counties sampled from D^0 will serve as pure control and will not receive any baraza at any level $\left(S_0^0\right)$. About one quarter will receive a sub-county level baraza that combines both information and deliberation treatment $\left(S_{ID}^0\right)$. One quarter will receive a sub-county level baraza that consists of officials providing information and limited opportunity for citizens to engage $\left(S_I^0\right)$. One quarter will receive a sub-county level baraza with a focus on citizens engaging with each other and with officials, with limited information provision $\left(S_D^0\right)$. We also take a random sample of sub-counties from the D^{ID} districts that received the district level baraza $\left(S_0^{ID}\right)$.

[1] The fact we restrict ourselves to eligible districts and sub-counties in this way has obvious consequences for the study population. For instance, the three districts that together constitute Karamoja, one of the poorest and most remote areas located in the northeast of Uganda, had already received district level barazas. As a result, this area was excluded from the analysis. While the impact evaluation had national coverage, this fact that we were evaluating and ongoing policy project needs to be kept in mind when interpreting results and extrapolating conclusions.

[2] As will be explained in detail below, we do not differentiate between the information component and the deliberation component of barazas organized at this level. In other words, district level barazas have both an information and deliberation component, and are thus the district level equivalent of sub-county level barazas that combines both information and deliberation treatment.

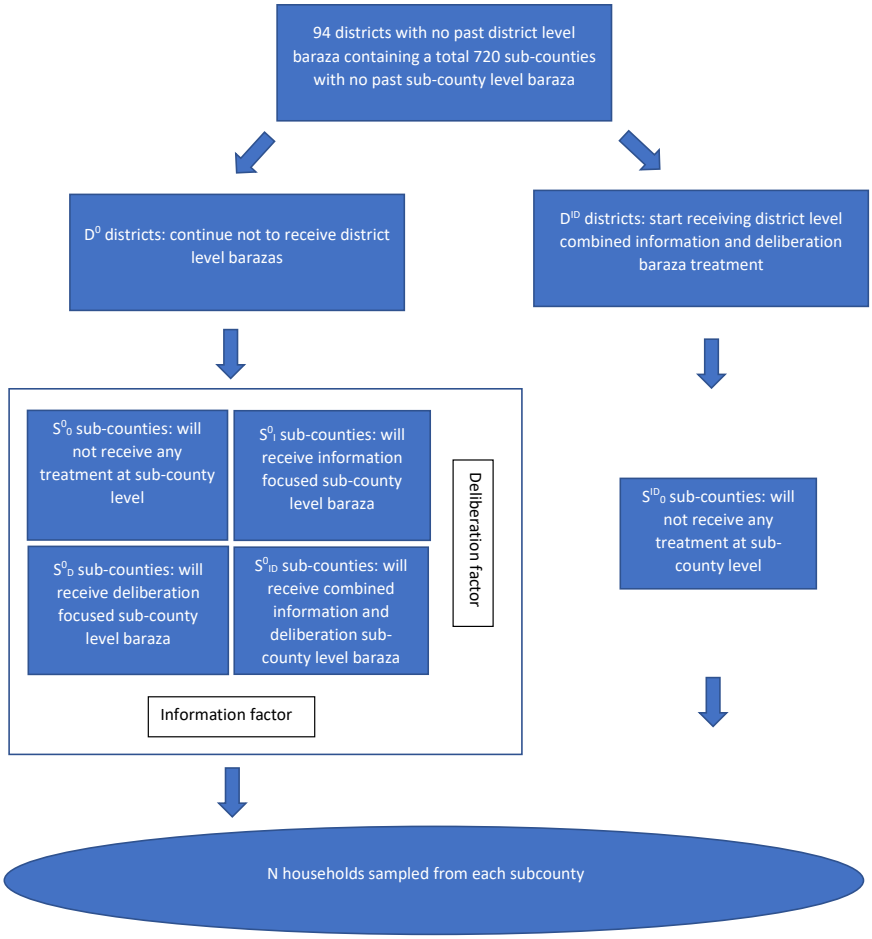

Figure 1: Experimental Design

Within each sub-county, we sample a fixed number of households.

The above design allows us to answer the four research questions in Section 3. First, to assess the impact of the sub-county baraza interventions as implemented by the government of Uganda, one can compare outcomes of households that were sampled from S_{ID}^0 to outcomes of households that were sampled from S_0^0. Second, to assess the relative importance of the information component of a baraza, one can compare outcomes of all households that were exposed to the information component (either as a stand alone information baraza as implemented in S_I^0 or as part of a combined baraza as implemented in S_{ID}^0) to outcomes of all households that were not exposed to the information component of the baraza (either because they did not receive a baraza at all $\left(S_0^0\right)$ or because they only received a deliberation focused baraza $\left(S_D^0\right)$). Similarly, to assess the relative importance of the deliberation component of a baraza, one can compare outcomes of all households that were exposed to the deliberation component (either as a stand alone deliberation baraza as implemented in S_D^0 or as part of a combined baraza as implemented in S_{ID}^0) to outcomes of all households that were not exposed to the information component of the baraza (either because they did not receive a baraza at all $\left(S_0^0\right)$ or because they only received an information baraza $\left(S_I^0\right)$). Note that, due to the factorial design, much more information can be used to test the second and third hypotheses than the first hypothesis. Finally, to investigate administrative placement of the intervention, two comparisons will be used. First, and similar to the first hypothesis, we can simply estimate the impact of district level barazas by comparing outcomes of households that were sampled from S_0^{ID} to households that were sampled from S_0^0. However, we can also directly compare district level barazas to sub-county level barazas by comparing outcomes of households that were sampled from S_0^{ID} to outcomes of household that were sample from S_{ID}^0[3].

To determine the number of districts, sub-counties and households to include in the study, we ran an extensive series of power calculations that used data from the Uganda National Household Survey of 2009/10 and the Demographic and Health Survey of 2011 to estimate standard errors of the outcomes and inter-class correlations. Outcomes used to determine sample size included weight-for-age z-scores for children; number of days unable to work as percentage of days sick at the household level; number of years the average child within the household goes to school, proportion of children in the household currently attending school; the proportion of households that was visited by an extension worker in the previous year; maize yields; time to get drinking water (including waiting time); and share of households having access to improved drinking water sources. This resulted in the selection of a total sample size of 11,500 households

[3]From a policy perspective, in light of the shift from sub-county level barazas to district level barazas, the last comparison is the most interesting one, and this comparison was indeed also the one that was in the pre-analysis plan. However, the partial roll-out of the intervention means that for this comparison we are constrained by the number of sub-counties in S_{ID}^0 that ended up being treated. Comparisons of outcomes in areas that received a district level baraza treatment to areas that did not receive a baraza have higher statistical power because more observations can be used.

distributed over 230 sub-counties in 40 districts throughout Uganda, on which baseline data was collected.[4] More details on the power calculations can be found in an online appendix.

In Table 1, we test for balance between the treatment groups at baseline. Baseline data on 12,545 households was collected between June and September 2015. Sample averages are reported in the first column (with standard errors in brackets below). For example, we see that the average household consists of about 6 household members and that only about 11 percent of households report that they have access to agricultural extension. In the second column, we report differences between baseline characteristics of households that were planned to receive a sub-county level combined information and deliberation baraza, and those that would not be exposed to any baraza. We find that distance to the nearest road is slightly though statistically significantly higher in areas that were planned to receive a sub-county level baraza. When comparing households that were planned to be exposed to a sub-county level information baraza to households that would not receive a sub-county level information baraza (column 3), we see that that households are slightly larger in the former group, and the difference is significant. The average household has two to three children attending a public school. We also find a slight pre-treatment imbalance on this outcome for the information treatment.

In the fourth column of Table 1, we report differences between households that would be exposed to a sub-county deliberation baraza and households that would not. For this treatment, we cannot reject balance on any of the variables. In the last column, we report differences in outcomes between households that would be exposed to a district level baraza and households that would not be exposed to any baraza. Also here, we do not find any significant difference. Overall, out of 40 comparisons, we find that two differences are significant, which is what one would expect to find due to chance alone. As such, we conclude that the randomization of the original design was successful.

5 Treatments

We tried to stay as close as possible to the baraza intervention as designed and implemented by the government before the start of the study. A description of a generic sub-county level baraza was already given in Section 2. We used this baraza as a starting point and, either removed the information component or the deliberation component to test the relative importance of these components. To standardize the treatments, we developed detailed scripts that RDCs and facilitators were expected to follow. Furthermore, manuals for RDCs and facilitators were developed, and all were invited for a training. We will summarize the main differences between an information baraza and a deliberation baraza. Detailed information can be found in an online appendix.

[4] We added an additional 3 sub-counties in each of the five treatment groups to account for attrition.

Table 1: Orthogonality tests

	mean	sc baraza	information	deliberation	level
Household size	6.324	0.021	0.304*	-0.003	0.267
	(2.825)	(0.142)	(0.133)	(0.125)	(0.226)
Age of the household head (years)	46.501	0.736	0.464	0.725	1.827
	(14.615)	(0.681)	(0.594)	(0.714)	(0.943)
Head of household is woman (1=yes)	0.191	0.012	-0.014	0.004	-0.006
	(0.393)	(0.014)	(0.013)	(0.015)	(0.012)
Head finished primary education (1=yes)	0.213	-0.007	-0.020	-0.003	-0.038
	(0.410)	(0.017)	(0.020)	(0.020)	(0.028)
Thatched grass roof (1=yes)	0.298	-0.001	0.009	-0.032	0.020
	(0.457)	(0.026)	(0.025)	(0.023)	(0.034)
Traditional mud wall (1=yes)	0.424	0.021	-0.025	0.038	-0.018
	(0.494)	(0.043)	(0.040)	(0.039)	(0.107)
Distance to nearest all weather road (km)	0.906	0.167*	0.106	0.147	-0.043
	(0.915)	(0.106)	(0.095)	(0.092)	(0.091)
Access to extension (1=yes)	0.108	0.002	0.004	0.007	0.007
	(0.310)	(0.014)	(0.012)	(0.014)	(0.019)
Village Health Team in village (1=yes)	0.854	0.000	0.006	0.025	0.079
	(0.353)	(0.031)	(0.026)	(0.026)	(0.034)
Number of children in public schools	2.478	0.044	0.165+	0.038	0.165
	(2.074)	(0.095)	(0.091)	(0.089)	(0.146)
Number of observations	12545	5193	10241	10241	4850

Note: First column reports sample means (and standard deviations below); Column 2 reports differences (and standard errors below) between baseline characteristics of households that were planned to receive a sub-county level combined information and deliberation baraza, and those that would not be exposed to any baraza; Column 3 reports differences (and standard errors below) between baseline characteristics of households that would receive a sub-county level information baraza, and those that would not receive a sub-county level information baraza; Column 4 reports differences (and standard errors below) between baseline characteristics of households that were allocted to receive a sub-county level deliberation baraza, and those that would not receive a sub-county level deliberation baraza; Column 5 (and standard errors below) reports differences between baseline characteristics of households that were planned to receive a district level combined information and deliberation baraza, and those that would not be exposed to any baraza; **, * and + denotes significance at the 1, 5 and 10 percent levels. Reported standard errors are clustered at the level of randomization. Distance to all weather road is trimmed at 5 percent and transformed using the inverse hyperbolic sine transformation.

For information (and combined information and deliberation) barazas, templates to gather information were developed to be filled by officials and mounted at a central location in each parish of the sub-county (or the district in case of a district level baraza) two weeks before the baraza. The template was designed to inform citizens about planned and actual public expenditures for the previous fiscal year, about achievements and challenges encountered during that year, and about planned expenditures and targets for the next fiscal year. This needed to be filled for each of the four sectors (agriculture, infrastructure, health and education) by the sub-county chief.[5] On the day of the baraza, the CAO provides a brief presentation on the overall budget for the fiscal year, the main achievements and challenges in service delivery, and introduces local officials. After a brief intervention by the OPM, local officials responsible for each sector then present more or less the same as what was required for the templates. An information-focused baraza allowed for only 10 clarifying questions to be asked, to be collected and asked by the facilitator.

For the deliberation barazas, posters are also mounted in each parish of the sub-county, but only to announce that a baraza will be held at a particular date and place. At the baraza itself, after a brief introduction by the OPM, citizens are guided to break into 4 groups by sector, discuss problems they face and draw up a list of priority issues that need to be addressed. Facilitators in each group are required to anonymously collect these issues and concerns. Facilitators are expected to focus the discussion on what was done well, and on problems encountered during the past year. The discussions should also result in agreement on what should be done in the next fiscal year. After the break-out sessions, officials are asked to react to the specific comments and requests.

A sub-county level baraza in the crossed treatment group simply combined both elements. District level baraza were very similar to sub-county level barazas, except for the fact that district level barazas are organized at the district headquarters, all sub-county chiefs and LC3's of each sub-county are expected to attend in case questions arise related to their sub-county, and all citizens within the entire district are invited.

6 Implementation challenges

The implementer of the baraza project, the OPM, experienced significant challenges in the roll-out of the intervention, affecting sub-county level barazas in particular.[6] At the start of 2018, and almost two and a half years after baseline data was collected, only about 25 percent of the planned sub-county level interventions had happened. At that point, costs and benefits of waiting until OPM finished all sub-county barazas or collecting baseline information after

[5] While the preparation and distribution of these posters were the responsibility of the RDC, we had research assistants that closely monitored the implementation of this. We also assisted in printing of the templates.

[6] All district level barazas were implemented as planned, so we mainly focus on sub-county level barazas here.

incomplete roll-out were assessed and various scenarios and strategies were evaluated[7]. After an additional six months, with still only 56 out of the 155 sub-county barazas implemented, it appeared that the best scenario would be one whereby end-line data would be collected before all sub-counties were treated, given that the governments financial constraints made it likely that complete roll-out would take place only very slowly if at all.

End-line data collection after partial roll-out may introduce selection bias. It may be that, from the randomly assigned treatment sub-counties, particular sub-counties were selected to be treated first and others postponed. For instance, for logistical reasons, the implementing partner may have started with sub-counties that were close to the capital, resulting in a treatment group that differs from the control group.

In this section, we reflect on various aspects related to this partial roll-out and report on analyses run and decisions made prior to endline data collection (and included in the pre-registered report, see below) to diagnose and accommodate potential selection bias. We start with a diagnosis of the problem through the comparison of baseline characteristics between treatment sub-counties that ended up not being treated and control sub-counties. We then explain an end-line data collection strategy that is likely to reduce selection bias in the event that the partial roll-out introduced selection bias. We also provide updated power calculations, as the reduced sample size will have altered the size of the effects we can confidently detect.

6.1 Balance between planned-to-treat-but-not-treated sub-counties, and control sub-counties

We can investigate if selection bias was introduced by the partial roll-out by comparing outcomes in control sub-counties to outcomes in sub-counties that were allocated to receive treatment but did not end up receiving treatment. The idea is that if the roll-out was random, sub-counties that were allocated randomly to a particular treatment at the design stage but did not end up receiving treatment can be interchanged with sub-counties that were randomly selected at design stage to function as control sub-counties. Finding no significant differences in outcomes between these two groups would support the hypothesis that the partial roll-out did not introduce selection bias. If the incomplete roll-out introduced selection bias, comparing these two groups may also be informative to assess the direction and magnitude of the bias. Note that such comparisons can be done using baseline data, but may be particularly informative after endline data collection.

Table 2 presents the original balance table (Table 1), but after dropping sub-counties that were treated. We find that the imbalance related to distance to nearest all weather road found in Table 1 becomes even more pronounced after dropping sub-counties that ended up being treated. This may indicate that

[7] One such strategy was a staggered endline data collections strategy, whereby endline data would be collected in areas where a baraza took place early in the project while we waited for the remaining barazas to be rolled out.

	Control	Information
Control	Planned: 51 Included: 40	Planned: 51 Treated: 29
Deliberation	Planned: 51 Treated: 18	Planned: 53 Treated: 20

Figure 2: Factorial design

OPM prioritized treatment of less remote areas. Similarly, for the information baraza, the imbalance for the number of children in school becomes stronger, suggesting information barazas were organized in areas where fewer children went to school to begin with. At the same time, it should be noted that significant differences are only found for comparison on which balance was rejected in the original balance table. As such, the incomplete roll-out may have actually improved baseline balance. Similar comparisons using endline data are reported in Section 9.

6.2 Selection of control sub-counties to be included in endline survey

Figure 2 summarizes the factorial design that underlies the sub-county level baraza impact evaluation. It shows that, based on the original power calculations, we planned to have between 51 and 53 in each treatment cell. At the start of the endline data collection, 29 sub-counties of the 51 that were supposed to receive the information-only treatment were treated, 18 sub-counties of the 51 that were supposed to receive the deliberation-only treatment were treated, and 20 sub-counties of the 53 that were supposed to receive the crossed (information + deliberation) treatment were treated. All of the 8 districts that were supposed to be treated were treated, from which 40 sub-counties were sampled. Overall, this means that about 55 percent of the sub-counties that needed to receive any form of baraza, received the treatment.

As only part of the intervention was implemented, it may not be cost effective to collect end-line data on all sub-counties that did not receive a treatment (either because they were allocated to the control or because they ended up not being treated). Statistical power is optimal when the number of treatment units is equal to the number of control units, and while adding more control sub-counties will increase power, the gains in power have the be weighed against the cost of collecting additional endline data. We decided to collect endline data on 40 control sub-counties, which is equal to the number of sub-counties that was treated with a district level baraza. This raises the question: from the potential control sub-counties (either those that were allocated to the control or because they ended up not being treated), which control sub-counties should be included in the endline data collection? One reasonable suggestion would be to pick them

Table 2: Balance between planned but not treated subcounties and planned controls

	mean	sc baraza	information	deliberation
Household size	6.324	0.012	0.388*	0.022
	(2.825)	(0.171)	(0.170)	(0.140)
Age of the household head (years)	46.501	0.357	0.698	0.553
	(14.615)	(0.714)	(0.663)	(0.808)
Head of household is woman (1=yes)	0.191	0.008	-0.019	-0.003
	(0.393)	(0.017)	(0.016)	(0.017)
Head finished primary education (1=yes)	0.213	-0.007	-0.007	-0.003
	(0.410)	(0.019)	(0.027)	(0.022)
Thatched grass roof (1=yes)	0.298	-0.002	0.000	-0.036
	(0.457)	(0.029)	(0.024)	(0.027)
Traditional mud wall (1=yes)	0.424	0.007	-0.057	0.044
	(0.494)	(0.049)	(0.047)	(0.044)
Distance to nearest all weather road (km)	0.906	0.284**	0.010	0.187
	(0.915)	(0.131)	(0.100)	(0.110)
Access to extension (1=yes)	0.108	0.005	0.008	0.007
	(0.310)	(0.015)	(0.016)	(0.015)
Village Health Team in village (1=yes)	0.854	-0.007	-0.010	-0.015
	(0.353)	(0.035)	(0.028)	(0.028)
Number of children in public schools	2.478	0.043	0.249*	0.076
	(2.074)	(0.112)	(0.115)	(0.100)
Number of observations	12545	4293	7842	8391

Note: First column reports sample means (and standard deviations below); Column 2 reports differences (and standard errors below) between baseline characteristics of households that were planned to receive a sub-county level combined information and deliberation baraza but did not receive one, and those that were planned not receive any baraza; Column 3 (and standard errors below) reports differences between baseline characteristics of households that were planned to receive a sub-county level information baraza but did not receive one, and those that were planned to not receive a sub-county level information baraza; Column 4 reports differences (and standard errors below) between baseline characteristics of households that were panned to receive a sub-county level deliberation baraza but did not get one, and those that were planned to not receive a sub-county level deliberation baraza; **, * and + denotes significance at the 1, 5 and 10 percent levels. Reported standard errors are clustered at the level of randomization. Distance to all weather road is trimmed at 5 percent and transformed using the inverse hyperbolic sine transformation.

randomly. However, if the roll-out was not random, such a strategy may lead to a biased estimate of the causal impact of the intervention. For instance, in section 6.1 we find indications that OPM may have prioritized treatment of less remote areas. Randomly selecting control sub-counties may result in a sample in which sub-counties closer to roads are relatively under-represented and sub-counties that are in more remote areas relatively over-represented in the control group. A better strategy may be to match, ex-ante, each treated sub-county to a control sub-county that is similar in a range of observable pre-treatment characteristics that the planner had access to when rolling out the intervention and are likely to have affected his or her decision (Kasy, 2016; Bertsimas, Johnson, and Kallus, 2015). For instance, a sub-county that has been treated may be matched to a candidate untreated sub-county on the basis of road network density.

We used a range of sub-county characteristics that were likely to be known to the planner and may have affected how the intervention was rolled out to match each treated sub-county to a control sub-county that was similar in terms of these characteristics. More in particular, we match on the following characteristics that were obtained at baseline from a survey of village chairs and chief administrative officers (CAO) of each sub-county: GPS coordinates of the sub-county, road infrastructure within the sub-county (km tarmac road and km all-weather (gravel) road), share of households with electricity, share of households with an iron roof or tiles, number of health centers in the sub-county, female primary school dropout rate, number of Universal Primary Education (UPE) schools in the sub-county, percent of farmers that use improved seed, and political connections of the sub-county (defined by having a minister or member of parliament coming from the sub-county). These characteristics are used in a probit regression to predict the probability that a sub-county would be treated. The probabilities are then used to match the each treated sub-county to a candidate control sub-county that is closest in terms of the probability of being treated.[8]

In Table 3, we look at baseline balance for the resulting sample. The imbalance that was found in Table 1 for the information treatment on household size and the number of children in school has disappeared. Consistent with the suspicion that OPM may have prioritized treatment of less remote areas, we now find that distance to nearest all weather roads is on average slightly higher in control sub-counties. Two significant results out of 40 comparison is again what can be expected from pure chance alone and so we conclude that also with this new sample we maintain balance between treatment and control on a range of baseline characteristics for the various hypotheses we will test.

While matching treatment and control sub-counties ex-ante may reduce bias

[8] In section 6.1, we diagnose potential selection bias introduced by the partial roll-out by comparing baseline characteristics of planned treatment sub-counties that ended up not being treated to that of control sub-counties. To enable similar comparisons using endline data, we need to make sure we have data on a sufficient number of planned treatment sub-counties that ended up not being treated to enable comparisons similar to those reported in Table 2. As a result, we implemented the matching of treatment and control sub-counties in a stratified way.

Table 3: Orthogonality tests for final sample

	mean	sc baraza	information	deliberation	level
Household size	6.411	-0.186	0.065	-0.302	0.062
	(2.855)	(0.169)	(0.152)	(0.166)	(0.248)
Age of the household head (years)	47.009	1.096	-0.215	0.574	1.554
	(14.542)	(1.012)	(0.731)	(1.038)	(0.998)
Head of household is woman (1=yes)	0.191	0.025	-0.006	0.022	0.011
	(0.393)	(0.017)	(0.018)	(0.024)	(0.015)
Head finished primary education (1=yes)	0.208	0.005	-0.016	0.014	-0.018
	(0.406)	(0.029)	(0.025)	(0.035)	(0.031)
Thatched grass roof (1=yes)	0.262	0.015	0.044	-0.007	0.037
	(0.440)	(0.030)	(0.030)	(0.022)	(0.042)
Traditional mud wall (1=yes)	0.444	0.086	0.031	0.062	-0.008
	(0.497)	(0.058)	(0.053)	(0.058)	(0.114)
Distance to nearest all weather road (km)	0.909	-0.279*	0.027	-0.104	-0.229
	(0.912)	(0.136)	(0.140)	(0.135)	(0.112)
Access to extension (1=yes)	0.105	0.011	0.000	0.012	0.018
	(0.307)	(0.014)	(0.012)	(0.020)	(0.016)
Village Health Team in village (1=yes)	0.865	0.020	0.019	0.090*	0.075
	(0.342)	(0.051)	(0.036)	(0.039)	(0.041)
Number of children in public schools	2.507	-0.089	0.001	-0.188	0.078
	(2.072)	(0.118)	(0.097)	(0.111)	(0.154)
Number of observations	7340	2949	5298	5298	3999

Note: First column reports sample means (and standard deviations below); Column 2 reports differences (and standard errors below) between baseline characteristics of households that received a sub-county level combined information and deliberation baraza, and those that did not receive any baraza; Column 3 reports differences (and standard errors below) between baseline characteristics of households that received a sub-county level information baraza, and those that did not receive a sub-county level information baraza; Column 4 reports differences (and standard errors below) between baseline characteristics of households that received a sub-county level deliberation baraza, and those that did not receive a sub-county level deliberation baraza; Column 5 reports differences (and standard errors below) between baseline characteristics of households that received a district level combined information and deliberation baraza, and those that did not receive any baraza; **, * and + denotes significance at the 1, 5 and 10 percent levels. Reported standard errors are clustered at the level of randomization. Distance to all weather road is trimmed at 5 percent and transformed using the inverse hyperbolic sine transformation.

resulting from incomplete roll-out, there are also costs involved. First, if sample selection is introduced by the roll-out, matching may further reduce external validity of the study, as now also the control sub-counties are not a random sample of the study population anymore. Second, the reduction in potential bias for hypotheses related to the sub-county level barazas should be traded off against an increase in potential bias when testing differences between control and district level barazas. As the sub-county level analysis weighed higher in terms of research objectives, we decided to prioritize the reduction of bias resulting from incomplete roll-out at this level. However, it should be kept in mind that both of these issues only become relevant if significant selection bias was introduced through the partial roll-out. Judged from the different balance tables using baseline data, this does not appear to be the case.

6.3 Updated power calculations

The original power calculations assumed full roll-out of the intervention. The partial roll-out necessitated an update to these power calculations to obtain a new set of minimal detectable effects (MDEs) associated with the sample that would be collected at endline (after partial roll-out and matching 40 control sub-counties to treated sub-counties). Below, we use baseline data to simulate MDEs for a selection of the outcomes we will use to judge effectiveness of the intervention (and are described in detail in section 7). We use a standard significance level of 0.05 (double sided). MDEs are estimated using a simple analysis of covariance (ANCOVA) model that controls for the outcome at baseline.

Figure 3 plots MDEs against power for the first outcome variable that will be used to assess the impact of barazas on public service delivery in the agricultural sector (extension at home, measured as the percentage of households in our sample who report that they were visited by an expert in the previous year). On average, about 11 percent of households in our sample reported that they were visited by an extension officer in the year before baseline data collection. The gray solid line shows the power curve associated with the deliberation treatment, comparing about 1,900 households that received the information treatment to about 3,450 households that did not receive a deliberation focused baraza. The light blue dashed line closely tracks the gray line, and shows power for different MDEs for the information component of the baraza intervention. Here, we compare about 2,450 households that live in sub-counties that received an information baraza to about 2,900 households that did not receive a sub-county information baraza. We also investigate power for the comparison between pure control barazas and sub-county level barazas (black dotted line). Here we compare about 1,000 households that received the combined information and deliberation sub-county level baraza to about 2,000 households that did not receive any baraza. Finally, the dark blue dashed line plots power for the comparison between about 2,000 pure control barazas and about 2,000 households that were exposed to a district level baraza.

Not surprisingly, we have most power for testing the information treatment. We see that the power curve hits the 80 percent threshold for the first time at an

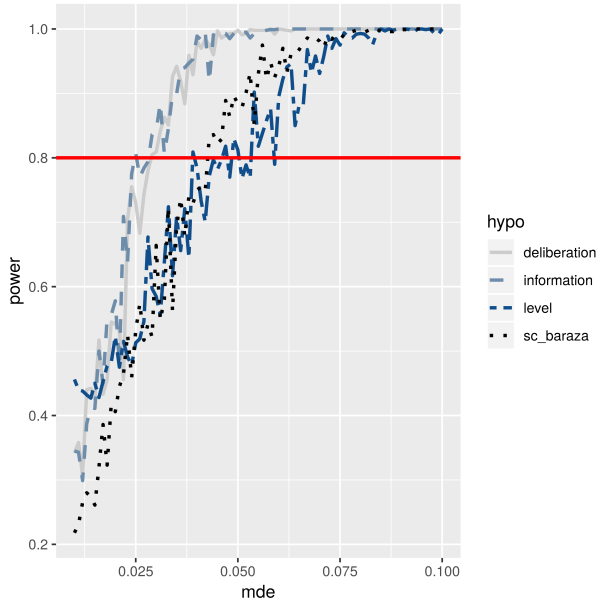

Figure 3: Power curves for access to extension

MDE of about 2.5 percentage points. The deliberation experiment is similarly powered, and at 80 percent we can expect to identify effects of 3 percentage points or more. Due to the smaller sample size, it is harder to compare pure control sub-counties to sub-county barazas that received the interacted treatment. Still, we are able to comfortably identify an increase of 4 percentage points. Despite a larger sample size available to test the impact of district level barazas, we find power to be comparable to that of the sub-county level baraza. This is due to the clustering at a much more aggregated level.

In figure 4, we plot MDEs for an infrastructure related outcome: distance (in km, trimmed at 5 percent and transformed using the inverse hyperbolic sine transformation) to the primary water source during the dry season. We find that for the information treatment and the deliberation treatment, we can detect a 4 percent difference at the standard 80 percent power level. As the average household lives about 900 meters from the primary water source, this translates in an average treatment effect of about 36 meters. Also here, the MDEs are higher when assessing the impact of the sub-county level barazas or the district level barazas. Then, MDEs correspond to about 60 meters for the average household in our sample. In an online appendix, we show results for similar analyses for all pre-registered variables that we will use to judge impact of the baraza intervention.

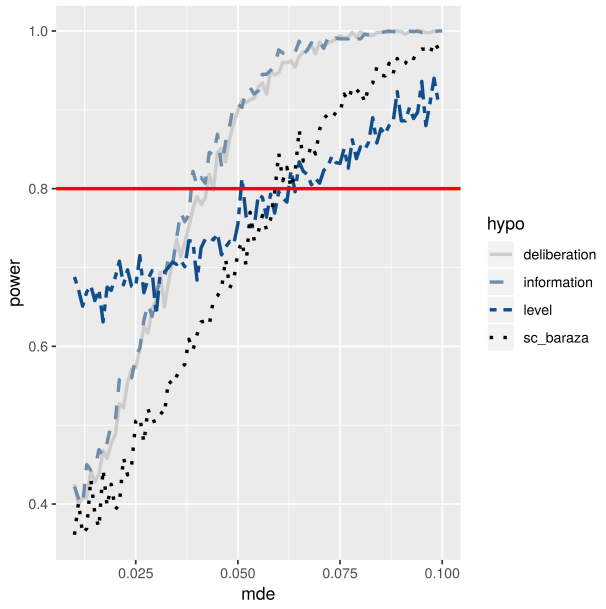

Figure 4: Power curves for distance to water source

6.4 Attrition

As more than four years passed between baseline and endline data collection, some level of attrition was to be expected. Overall, attrition amounted to 8.6 percent. Reports from the field indicated that this attrition was due to some villages close to the lake shores that were flooded due to heavy rains. Two villages in areas populated with land insecure settlers were dropped due to security concerns. Finally, attrition was generally higher in urban areas, as households that were residing in rented properties returned to their villages. Table 4 shows that attrition is independent of the various interventions of the study.

7 Results

We start by presenting results of a strictly confirmatory analysis for the four main hypotheses outlined in Section 3. Before end-line data collection, we prepared a "mock report", which contains the complete analysis on simulated end-line data for a core set of primary outcomes. This mock report was registered at the American Economic Association's RCT registry with a time stamp[9]. The

[9]The document uploaded in January 2020 to the AEA RCT registry is the first version of the pre-registered report and can also be found on github. However, later a coding mistake was found and corrected: For the comparison of outcomes between households that received a sub-county level barazas and households in the control group, we were reporting the inter-

Table 4: Attrition

	mean	sc baraza	information	deliberation	level
Attrition	0.086	0.010	0.0139	0.003	-0.004
	(0.280)	(0.020)	(0.017)	(0.019)	(0.014)
Number of observations	7340	2996	5341	5341	3996

Note: First column reports mean attrition rate (and standard deviations below); Column 2 reports correlation of attrition (and standard errors below) to the sub-county level baraza treatment; Column 3 reports correlation of attrition to the information baraza treatment; Column 4 reports correlation of attrition to the deliberation baraza treatment; Column 5 reports correlation of attrition to the district level baraza treatment. Reported standard errors are clustered at the level of randomization.

outcomes can be categorized into four broad sectors: agriculture, health, education, and infrastructure (drinking water and roads). With this report, we thus committed to key outcomes in each sector, and combine the outcomes in the four sectors into sector level indices, and in a single index. Pre-registration and mock reports have been suggested to reduce intentional or unintentional selection of specifications and outcome measures that yield positive findings, leading to an unreliable body of published research. (Humphreys, De la Sierra, and Van der Windt, 2013).[10]

Impact is assessed as a simple treatment-control comparison, implemented using an ANCOVA model that also controls for the region (as this was used for stratification) and the baseline outcome. In each specification, we also include all interaction terms of the factorial design (Muralidharan, Romero, and Wüthrich, 2019). Standard errors are clustered at the level of randomization: at the sub-county level for the first three hypotheses and at the district level for the last hypothesis.

Figure 5 provides a summary of the results. It shows the impact of the four main hypotheses—the impact of the sub-county baraza, the relative effectiveness of the information component, the relative effectiveness of the deliberation component, and the impact of district level barazas—on four sectors we consider—agriculture, infrastructure, health, and education. The graphs are based on indices that are composed of individual outcomes in each sector and aggregated following Anderson (2008); point estimates are standardized effects and confidence intervals are obtained following the permutation method explained

action effect between information and deliberation instead of the combined information and deliberation effect. This was corrected on April 2, 2020 in this commit (9f5afdbfdd6be766).

[10] To further increase transparency and allow for easy replication, the document was prepared using Lyx, an open source Latex front-end and all Latex and R code to replicate the analysis is placed under revision control using Git. The Git repository can be found at https://github.com/bjvca/baraza/.

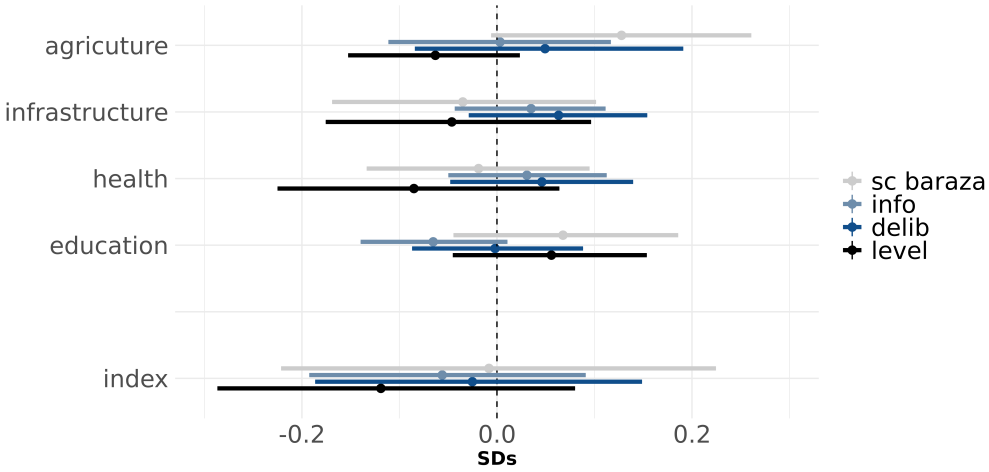

Figure 5: Summary of baraza impact

in Gerber and Green (2012).[11] We also combine the four indices into one overall index that assesses the impact on public service delivery in general. We find no significant impact of the baraza programme on overall public service delivery. There are some indications that sub-county level baraza did make a difference in the agricultural sector.

The indices combine various outcomes, some of which the expected direction of the effect is unclear a-priori. For instance, an information baraza may increase the quality of services in a hospital or health centre when judged by an objective measure such as waiting time. However, the information may also result in higher expectations from the part of the user. As such, perceptions of quality may have reduced as a result of an information baraza. It is therefore also interesting to look beyond the indices and consider outcomes individually.

7.1 Agriculture

We first zoom in on the outcomes that are used to assess the effectiveness of barazas in changing service delivery an associated outcomes in agriculture. Results are reported in Table 5. We start by looking if the baraza programme affected the use of modern inputs in agriculture. A first outcome looks at whether the household used inorganic fertilizers (DAP, Urea, NPK, Foliar, TSP, SSP,

[11] This method first reconstructs a complete schedule of potential outcomes by adding and subtracting the average treatment effect for control and treated units respectively. These potential outcomes are then used to simulate all possible random allocations. For each allocation, average treatment effects are estimated and 2.5th and 97.5th percentiles are then taken as the lower and upper limits of the 95 percent confidence interval. As this provides a conservative estimate of the confidence interval, it may be that results in the graph appear insignificant, but do show up as significant when testing significance using Fisher's exact test.

MOP) during the last 12 months. The first column reports sample averages, with standard deviation in brackets below. We find that about 23 percent of households in the sample give an affirmative answer to this question. In the second column, we report differences in outcomes between households that received a sub-county level baraza (i.e. the crossed treatment of a sub-county information baraza and a sub-county deliberation baraza; the bottom right in Figure 2) and households that did not receive any baraza (pure control; the top left in Figure 2). We see that the proportion of households that reports using inorganic fertilizer is 1.5 percentage points lower among the sub-group of households that were exposed to a sub-county level baraza that consists of both the information and the deliberation component than among households that did not receive any baraza (second column). However, this difference is not statistically significant. In the third column, we report differences between outcomes of households that live in areas where an information baraza was organized (either only an information baraza or a crossed information and deliberation baraza; top and bottom right of Figure 2) and outcomes of households that live in areas that were not exposed to an information baraza (either pure control or only a deliberation baraza; top and bottom left of Figure 2). We see that adoption of inorganic fertilizer is 3.4 percentage points higher among households that were exposed to an information baraza. However, also here, the difference is not significant. In the fourth column, we report differences between outcomes of households that live in areas where a deliberation baraza was organized (either only a deliberation baraza or a crossed information and deliberation baraza; bottom left and right of Figure 2) and outcomes of households that live in areas that were not exposed to a deliberation baraza (either pure control or information only baraza; top left and right of Figure 2). We also do not find differences in terms of inorganic fertilizer use. Finally, in the fifth column, we compare households that were exposed to a district level baraza to pure control households. Again, no impact of the district level baraza is found on this outcome.

The second outcome is related the use of improved seed. This input seems to be used more widely than inorganic fertilizer: 36 percent of households report that they have been using improved seed during the last year. This percentage is 4.3 percentage points higher among households that reside in areas where a sub-county level baraza took place as opposed to in areas where no baraza was conducted, but the difference is not significant. We find negative point estimates for the relative effects of both the information and the deliberation component, but effects are imprecisely estimated. Finally, we find adoption of improved seed to be lower in areas where a district level baraza was conducted, but the difference with areas that did not receive a baraza is not significant. Adoption of improved seed and inorganic fertilizer was included in the index that is used for the confirmatory analysis.

Next, we find that about 12 percent of households report that they received improved seed from the government extension system (through an extension agent, from the National Agricultural Advisory Services (NAADS) or through Operation Wealth Creation (OWC) that replaced NAADS). We find that this is 5.1 percentage points higher in areas where a sub-county level baraza took

Table 5: Treatment-control differences (ANCOVA) - Agriculture

	mean	sc baraza	information	deliberation	level
Used inorganic fertilizer (yes/no)†	0.229	-0.015	0.034	0.001	-0.013
	(0.420)	(0.033)	(0.035)	(0.049)	(0.031)
Used improved seed (yes/no)†	0.364	0.043	-0.030	-0.037	-0.043
	(0.481)	(0.033)	(0.038)	(0.038)	(0.034)
Obtained seed from government (yes/no)	0.121	0.051^{+}	0.004	0.056^{+}	-0.005
	(0.326)	(0.024)	(0.025)	(0.043)	(0.015)
Used agro-chemical inputs (yes/no)	0.469	-0.028	-0.007	-0.005	-0.010
	(0.499)	(0.050)	(0.035)	(0.046)	(0.043)
Used improved livestock methods (yes/no)	0.221	0.029	0.021	0.030	-0.014
	(0.415)	(0.031)	(0.028)	(0.034)	(0.026)
Extension visist at home (yes/no)†	0.178	0.056^{+}	0.037	0.036	-0.027
	(0.383)	(0.018)	(0.030)	(0.048)	(0.014)
Visisted extension office, demo site or model farmer (yes/no)†	0.285	0.040	0.036	0.045	-0.013
	(0.452)	(0.028)	(0.035)	(0.044)	(0.023)
Need extension (yes/no)	0.769	-0.015	-0.043^{+}	-0.029	-0.038
	(0.421)	(0.024)	(0.025)	(0.026)	(0.017)
Extension agents aware of this need? (yes/no)	0.264	-0.006	0.017	-0.001	-0.075*
	(0.441)	(0.024)	(0.027)	(0.035)	(0.023)
Extension agents decided alone? (yes/no)	0.316	0.034	-0.041	-0.031	-0.056
	(0.465)	(0.033)	(0.027)	(0.034)	(0.033)
Are there any farmer associations/groups in this village? (yes/no)	0.403	0.060^{+}	0.040	0.087*	-0.032
	(0.491)	(0.030)	(0.038)	(0.041)	(0.027)
NAADS/OWC in village? (yes/no)†	0.173	0.070*	-0.015	0.053	-0.037
	(0.378)	(0.028)	(0.030)	(0.040)	(0.022)
Support in marketing from marketing committee (yes/no)†	0.069	0.018	-0.013	0.016	-0.014
	(0.254)	(0.022)	(0.017)	(0.017)	(0.012)
Support in marketing from cooperative (yes/no)†	0.062	0.037^{+}	-0.021	-0.001	-0.009
	(0.241)	(0.024)	(0.016)	(0.021)	(0.014)
Number of observations	6704	2738	4858	4858	3687

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports the effect (and standard errors below) of the administrative placement of the baraza intervention. **, * and + denotes significance at the 1, 5 and 10 percent levels as assessed using randomization inference. Reported standard errors are clustered at the level of randomization. † indicates that the outcome was in included in the index for the confirmatory analysis.

place. We further see that this is especially due to the deliberation component of sub-county level barazas. In areas where such a baraza was held, the share is 5.6 percentage points higher than in areas were citizens were not given the opportunity to engage with their leaders. We find no such effects from district level barazas. Direct comparison of district and sub-county level barazas indicate that sub-county level barazas are significantly more effective in increasing the likelihood that households report to have received these inputs from government.

We then check if household changed with respect to the use of agro-chemicals. This includes the use of pesticides, herbicides, fungicides and acaricides during the last 12 months. Overall, almost half of all households in the sample report using some form of agro-chemicals. We do not find evidence that the baraza intervention affected the use of these inputs. Finally, we consider the use of modern inputs and methods in livestock rearing over the last 12 months. This includes improved animal breeds, the use of modern feeds, drugs, and artificial insemination. 22 percent of households report that they used such inputs and this proportion is similar across different experimental groups.

We then turn to agricultural advisory services. We first investigate if the barazas have affected access to extension at home. We estimate the percentage of households in our sample who report that they were visited by an expert (e.g. crop or livestock extension agent, or community-based facilitator or another experienced farmer) at the home in the last 12 months. We find that access to extension, although higher than at baseline, is still low, with only about 18 percent of households reporting that they received such a visit. We find that this percentage is significantly higher among households that were exposed to a sub-county level baraza. The effect is large, amounting to an increase of 5.6 percentage points. The effect seems to come from a combination of the information and deliberation components; the components in itself do not seem to affect the outcome enough to render if significant. We also find that this effect is absent among households that live in sub-counties that received a the district level baraza. A direct comparison of extension at home between households that were exposed to a district level baraza and households that were exposed to a sub-county level baraza confirms that sub-county level barazas are significantly more effective.

Home visits by extension officers are not the only way in which households have access to information. Extension offices, demonstration sites and model farmers are also an integral part of the Ugandan agricultural advisory system. Especially after the establishment of NAADS, such a demand-led service component that can be consulted by farmers when the need arises became more important than the more supply-driven component of training and visit. We thus also inquire if anyone in the household visited an extension office, demonstration site, or model farmer in the past year. We find that about 29 percent of households in our sample report access to extension in this way. While the results are in line with extension visits at home, differences are not significant. Access to extension, both at home or though extension offices and demonstration sites, was also included as key outcomes to assess impact using in the agriculture

and overall index.

We find that three quarters of households in our sample mention that there are agricultural enterprises, improved technologies or inputs you would like to adopt, indicating significant demand for advisory services. We also find that, according to citizens, service providers and policy makers are not always aware of the needs of farmers. The table shows that only 26 percent of households is of the opinion that officials are aware of which services farmers need. While we do not see that this percentage differs between treatment and control for sub-county level barazas, we do see that a district level baraza reduced this percentage. This may indicate that a district level baraza makes the mismatch between what farmers need and what officials think farmers need more salient.

Related to the previous outcome, we ask how decisions related to what topics to cover in agricultural extension are made. We define this outcome in a negative way, that is, the indicator is true if decisions are made without consultation. We see that about 32 percent of households indicate that no consultation happens, and the content of extension advisory services is decided upon by experts at the central level. We do not find that the baraza intervention increased participation in extension service planning.

About 40 percent of households report the presence of farmer groups or cooperatives in their village. In the agricultural sector in Uganda, such groups are very important. They are actively promoted by the government. In fact, to be able to receive inputs from the government extension system, farmers are strongly encouraged to form such groups. We find that sub-county level barazas increase the likelihood that farmer cooperatives or groups are formed in the villages in Uganda. It seems that the deliberative component is the main driver behind this result. We also find that this effect is specific to interventions at the sub-county level. We further find that a higher share of farmer groups in areas that received a sub-county level baraza received support from government.

The final two questions focus more on marketing. Connecting farmers to markets is also an important strategy outlined in the Agriculture Sector Strategic Plan (ASSP). The first outcome relates to the likelihood that farmers are supported by government through a village procurement committee. In the sample, about 7 percent of households report that they were assisted by government. A second questions is similar but looks at the role of cooperatives in marketing support. We generally find no effects of the baraza intervention, except perhaps for an increase of almost 3.7 percentage points in the likelihood that cooperatives assist with marketing in areas that received a sub-county baraza. Both of these outcomes were also included in the index to assess overall impact.

7.2 Infrastructure

A second important area in which we expect to see an impact of the baraza programme is in infrastructure. We primarily focus on drinking water infrastructure. Results are in Table 6, which is organized similar to Table 5.

A first outcome we consider is whether the household uses an unprotected water source during dry season. This is measured as the share of households

that report that the main source of drinking water during the dry season is surface water, an unprotected dug well or an unprotected spring. We find that about 16 percent of households in the sample report that they are using an unprotected water source. The baraza intervention does not seem to affect this proportion. This outcome is included in the infrastructure index.

A second outcome we look at (and is also included in the index) is the distance to the primary water source during the dry season. This was measured in km, but trimmed and transformed using the inverse hyperbolic transformation. We find that, on average, households have to walk about 0.91 km. While this distance seems to reduce in all comparisons, and especially for barazas held at the district level where we find a reduction of approximately 9 percent, it is never significantly different from zero.

The third outcome, also part of the index, is the time that one must wait at the water source, measured in minutes. This continuous variable was also trimmed and transformed. We find that households must wait on average about 36 minutes. We find that waiting time in areas that were exposed to the sub-county level baraza intervention reduced by about 25 percent. The deliberation component seems to be mostly responsible for this reduction.

The fourth outcome variable assesses changes in the presence of a water user committee in the village. Overall, about 60 percent of households report that such a committee is present in their village. We do not find that this share varies between the different experimental groups. Consistent with the finding that deliberation has been found to increase subsequent citizen involvement in community affairs (Björkman Nyqvist, de Walque, and Svensson, 2017), we see that the deliberation component leads to an increase in the likelihood that a household member is a member of a water user committee. We do not find that the baraza intervention increases the likelihood that these committees hold public meetings.

Households were also asked if they were satisfied with the quality of the water that is available at the source during the dry season. About 62 percent respond that they are satisfied or very satisfied with the drinking water. We do not find that households that were exposed to the baraza intervention are more or less likely to report that they are (very) satisfied with the quality of drinking water during the dry season. Half of the households report that they treat drinking water before drinking it, either by boiling it or treating it with chlorine. The likelihood that households treat water reduces somewhat for the information treatment. This may be due to the belief held in many households that water from a safe sources does not need additional treatment.

We include one question related to road infrastructure. We ask how far the household is located from the nearest all weather road. We find that in the full sample a household lives on average 26 km from a road. We do not find that the baraza programme reduces the distance to the nearest all weather road.

Table 6: Treatment-control differences (ANCOVA) - Infrastructure

	mean	sc baraza	information	deliberation	level
Household uses unprotected water source during dry season (yes/no)†	0.159	0.031	0.005	0.010	-0.023
	(0.366)	(0.042)	(0.036)	(0.037)	(0.046)
Distance to water source (km)†	0.748	-0.026	-0.040	-0.049	-0.091
	(0.576)	(0.046)	(0.041)	(0.061)	(0.039)
Average waiting time at source (min)†	3.198	-0.286*	-0.006	-0.287+	-0.032
	(1.638)	(0.152)	(0.117)	(0.193)	(0.160)
Is there a Water User Committee in this village? (yes/no)†	0.598	-0.021	0.033	0.032	-0.009
	(0.490)	(0.046)	(0.037)	(0.040)	(0.047)
Satisfied with the qualiy of water? (yes/no)	0.624	0.031	-0.009	-0.062	0.020
	(0.484)	(0.052)	(0.044)	(0.044)	(0.038)
Boil/treat water before drinking (yes/no)	0.500	-0.025	-0.087*	-0.020	-0.050
	(0.500)	(0.045)	(0.037)	(0.046)	(0.056)
Member of commitee? (yes/no)	0.163	0.022	0.001	0.040+	0.002
	(0.370)	(0.021)	(0.017)	(0.025)	(0.019)
Committee holds public meetings? (yes/no)	0.474	-0.005	0.043	0.060	0.010
	(0.499)	(0.044)	(0.036)	(0.042)	(0.041)
Distance to nearest all weather road (km)†	2.849	0.388	-0.129	-0.286	0.591
	(1.788)	(0.314)	(0.306)	(0.313)	(0.405)
Number of observations	6704	2738	4858	4858	3687

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports the effect (and standard errors below) of the administrative placement of the baraza intervention. **, * and + denotes significance at the 1, 5 and 10 percent levels as assessed using randomization inference. Reported standard errors are clustered at the level of randomization. † indicates that the outcome was in included in the index for the confirmatory analysis. Distance to water source, average waiting time at source, and distance to nearest all weather road are all trimmed at 5 percent and transformed using the inverse hyperbolic sine transformation.

7.3 Health

We now look at outcomes in the health sphere (Table 7). One problem with public health related outcomes is that some outcomes related to quality of service delivery in health centres will only be available for households that have visited government health facilities, reducing sample size too much to maintain acceptable power. Furthermore, the intervention may also affect the likelihood that households visit a government health facilities, introducing selection bias.

The first two outcomes we consider attempt to assess changes in access or use of public health facilities. A first indicator measures demand for public health services for illness. In particular, we construct an indicator that is true if the household head responds that treatment would be sought in a health center 2, 3, 4 or in a regional referral hospital if a member of his/her household had fever. We find that 69 percent of households respond that they would seek treatment in a government health facility. This proportion is independent of the treatment groups.

A similar indicator attempts to assess demand for public health services for maternal health care, and asks if treatment would be sought in a health center 2, 3, 4 or in a regional referral hospital if a member of the household was to give birth. This percentage is even higher: about 81 percent would go to a government health facility to give birth. Again, this proportion is not affected by the baraza programme. Both outcomes are included in the health index.

Next, we ask if a Village Health Team (VHT) is present in the village. VHTs are very important in front-line health care provision in Uganda. They also have prominent roles in government health interventions, such as immunization campaigns or the distribution of bed nets. We find that overall, about 88 percent of households report that a VHT is present in their village. The presence of a VHT is not impacted by the baraza intervention.

As the baraza tries to increase citizen engagement, we also check if households that were exposed to a baraza are more likely to participate in VHTs. We thus asked if any member of the household was a member of a VHT. We see that in about 11 percent of our sample, at least one household member is part of a VHT. The baraza intervention does not increase the likelihood that individuals participate as VHT members. Furthermore, the baraza intervention attempts to encourages sharing of information. As such, we expect that being exposed to a baraza may encourage VHTs to organize more public meetings. We find that overall, 43 percent of households state that VHTs have organized a public meeting in the last year. We find that this proportion is significantly higher in areas that were exposed to a sub-county level baraza. This effect seems driven by the deliberative component of a sub-county baraza.

We also consider distance to the nearest government health facility, measured in km (trimmed and transformed using the inverse hyperbolic sine transformation). Overall, average distance to the nearest government health facility is almost 50 km. We do not find that barazas reduce this distance.

We then turn to health outcomes. We start by asking if any member of the household has been sick during the last year. This was the case in two thirds of

the households in our sample. The intervention did not reduce morbidity in our sample. We then ask for each sick person in the household to record how many days he or she was ill, and use this to calculate the total number of sick days at the household level in the last year. The average household recorded almost 50 sick days according to this definition. We also do not find that the intervention affected the (trimmed and transformed) number of sick days. We also look at the number of days household members were unable to go to school or to work, which provides an indication of the severity of illness. Calculated similarly to the previous outcome, we find that in the average household about 35 school- or workdays were missed due to illness. Again, there is no significant reduction in this (trimmed and transformed) number. This last health outcome measure was included in the health index.

We then ask how long one had to wait before being attended to (in minutes, trimmed and transformed). We find that the sample mean for this outcome is about 90 minutes. While we see that waiting time reduces for most comparisons, the differences are never significant. Potentially, the reduced sample size resulted in too little power to detect a difference. This outcome was also included in the health index.

The next question that was included in the index was again asked to all households. In particular, we inquire if a traditional health practitioner was consulted in the last year. In one in four households in our sample, this was the case. The baraza intervention did not affect this percentage.

One problem that often crops up in the health sector is absenteeism. To assess this, we ask who examined the patient in the health center. Ideally this should be the doctor or in-charge. If this person is absent, patients are generally examined by nurses or lab technicians. We thus construct an indicator that is one if the household responds that the patient was investigated by the doctor or the in-charge, and zero otherwise. Only in 41 percent of the cases, a qualified person appears to do the examination. Surprisingly, the deliberation component of the baraza seems to reduce this likelihood somewhat. We also look at the time that the examination takes. The average examination in our sample took about 22 minutes. There is no change related to the intervention.

Health care in Uganda is supposed to be free. However, corruption is widespread and often patients are required to make payments to receive care. We find that about 18 percent of households report that payment was required the last time they visited a government health facility. There is no impact of the intervention. Related, users often complain about a lack of drugs in government health facilities. We asked if, during the last visit to a government health centre, drugs were received (indicating that drugs were available). We also asked if drugs had to be purchased from outside of the hospital (indicating that at least some drugs were missing). While 71 percent of households report that they received medicines in the health centre, almost all of them also mentioned that they had to also buy drugs outside of the hospital. For neither indicator, the intervention seems to make a difference.

We further probe for a subjective assessment of the overall quality of care at the health facility. Most households report that they are satisfied or very

satisfied with services received at the government health facility. This seems to increase in areas where a sub-county level baraza took place and there is also a sizable difference in outcomes when comparing sub-county level baraza outcomes to district level baraza outcomes. However, none of the differences are significant.

We considered several other health related outcomes that feature prominently in other studies. One key outcome in Björkman and Svensson (2009) is immunization. However, we already find close to 100 percent immunization rates in our baseline data. Another outcome is child mortality. Child mortality rates at baseline were estimated at 38 per 1000 live births, which was deemed too low to include in the analysis. Raffler, Posner, and Parkerson (2018) find similar child mortality rates at baseline and suggest that the fact that they do not find an effect while Björkman and Svensson (2009) do is due to differences in baseline conditions: child mortality at baseline in Björkman and Svensson (2009) was 117 per 1000 live births.

7.4 Education

Education outcomes to assess impact of the intervention suffers from a similar problem as the one encountered with health outcomes: not all households in the sample have children in school, and so for many of the outcomes related to education, sample size becomes small. This also affects the indices. Results are presented in Table 8.

If the quality of public education is poor, households will be less likely to send their children to public schools. A first obvious outcome is thus to simply compare the number of children within the households that attend public school (either Universal Primary Education or Universal Secondary Education). We find that the average household in our sample had almost two children in government schools, but that enrollment rates are not affected by the baraza intervention.

Access to public education is also influenced by the distance to a public school. We thus recorded distance to primary or secondary school (or the average if both are reported; trimmed and transformed). We find that on average, households live about 3 km from a government operated school. Also for this outcome, the baraza program did not make an impact.

We also look at school infrastructure. First, we ask households if the primary or secondary school attended by any of their children has a complete boundary fence. In the complete sample, it was reported that only about 35 percent of schools have such a fence. The lack of a fence was a frequent complaints from parents during qualitative work. We also ask if the school has electricity and if there is a water source available in the school. We find that overall, about 34 percent of schools have electricity and about 70 percent have a water source. We find that sub-county level baraza seem to improve school infrastructure. We considered many other infrastructure related outcomes, such as the number of classrooms and availability of functioning toilets for both girls and boys, but baseline data suggested there were generally no issues related to these outcomes.

Table 7: Treatment-control differences (ANCOVA) - Health

	mean	sc baraza	information	deliberation	level
Seek treatment for fever in public health facility (yes/no)†	0.691	-0.008	-0.007	0.025	-0.010
	(0.462)	(0.033)	(0.033)	(0.040)	(0.046)
Go to public health facility to give birth (yes/no)†	0.813	-0.029	-0.033	-0.016	-0.070
	(0.390)	(0.034)	(0.029)	(0.035)	(0.043)
Is there a VHT in village? (yes/no)†	0.881	0.022	0.005	0.029	-0.019
	(0.323)	(0.031)	(0.025)	(0.025)	(0.027)
Are you or any member of the household part of the VHT (yes/no)	0.113	0.022	0.003	-0.001	-0.024
	(0.317)	(0.017)	(0.014)	(0.015)	(0.012)
VHT organised public meetings? (yes/no)	0.429	0.076*	-0.018	0.058	-0.046
	(0.495)	(0.041)	(0.033)	(0.040)	(0.036)
Distance to nearest govt health facility (km)†	3.875	0.256	-0.162	-0.252	0.445
	(1.377)	(0.219)	(0.233)	(0.263)	(0.342)
Any hh members sick? (yes/no)	0.658	0.003	0.024	0.037	-0.015
	(0.475)	(0.023)	(0.028)	(0.033)	(0.024)
Number of days sick?	2.576	-0.005	-0.040	0.004	-0.064
	(2.189)	(0.091)	(0.149)	(0.166)	(0.105)
Number of days work/school missed due to illness?†	2.273	-0.016	0.037	0.023	-0.065
	(2.027)	(0.093)	(0.134)	(0.153)	(0.121)
Waiting time before being attended (min)†	4.744	-0.040	-0.133	-0.151	0.064
	(1.012)	(0.093)	(0.108)	(0.135)	(0.082)
Has visited traditional health practitioner? (yes/no)†	0.257	-0.017	0.016	0.034	-0.039
	(0.437)	(0.032)	(0.029)	(0.030)	(0.019)
Was examined by in-charge/doctor (yes/no)	0.411	0.044	-0.049	-0.070+	-0.041
	(0.492)	(0.041)	(0.032)	(0.042)	(0.025)
Time of examination (min)	3.403	0.048	-0.099	0.015	-0.002
	(0.761)	(0.066)	(0.070)	(0.091)	(0.083)
Paid for services? (yes/no)	0.179	0.010	-0.008	-0.013	-0.005
	(0.384)	(0.023)	(0.025)	(0.042)	(0.024)
Received medication? (yes/no)	0.709	0.000	-0.003	0.000	-0.024
	(0.454)	(0.036)	(0.023)	(0.033)	(0.034)
Satisfied with services at health center (yes/no)	0.682	0.048	-0.026	-0.038	-0.011
	(0.466)	(0.033)	(0.031)	(0.038)	(0.026)
Number of observations	6704	2738	4858	4858	3687

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports the effect (and standard errors below) of the administrative placement of the baraza intervention. **, * and + denotes significance at the 1, 5 and 10 percent levels as assessed using randomization inference. Reported standard errors are clustered at the level of randomization. † indicates that the outcome was in included in the index for the confirmatory analysis. Distance to nearest govt health facility, number of days sick, number of days work/school missed due to illness, waiting time before being attended, and time of examination are all trimmed at 5 percent and transformed using the inverse hyperbolic sine transformation.

We also look at how the school is managed, and how parents are involved. For instance, we look at whether the school has a Parent Teacher Association (PTA) and a School Management Committee (SMC). Almost all schools have a PTA. We further find that 92 percent of households state the primary or secondary school attended by any of their children has a SMC. However, not all households are informed about SMC meetings. The baraza intervention does not seem to affect how schools are managed, how parents can participate, or how information is shared. Finally, we ask households if an inspector had visited the school in the year before the survey. We find that about 64 percent of households indicate that schools were inspected. Surprisingly, this proportion reduces as a result of the information component of a baraza.

7.5 Other outcomes

In this section, we look at some other outcomes, including some of the mechanisms through which the baraza intervention is assumed to impact outcomes. More in particular, we explore if the baraza program affected interfacing with politicians and civil servants, contributions to public goods (both cash and in-kind), and perceptions of service quality and prioritization.

One of the primary aims of the community forums is to increase communication between politicians, civil servants and the citizens. We thus try to assess if citizens interact more with politicians and service providers as a result of the interventions. We ask how long it has been since the respondent spoke personally to various officials for reasons related to service provision in agriculture, health, education, water or roads. Based on the answer, we construct an indicator variable that denotes if the household had a meeting or not. The time frame changes depending on the official. For instance, for the LC1 chairperson, the indicator takes the value of one if the respondent spoke to him or her within the last month. For the head teacher, the reference period is 6 months. For the other officials (sub-county chief, health management unit member and water committee member), the indicator is true if contact was sought in the past year. Results are in Table 9.

We find that about 43 percent of households in our sample have met with the LC1 chair in the month before the endline data was collected. The baraza intervention did not affect the likelihood that citizens meet with the LC 1 using this definition. About 20 percent of respondents report that they met with the sub-county chief in the last year. We do not find that the baraza intervention changes this likelihood. We do see that the information component of the baraza increases the likelihood that citizens interface with the head teacher or with members of the school management unit. We also see that the information component of the sub-county level baraza increases the likelihood of meetings with water committee members, but the difference is not significant. Finally, and similar to meetings with sub-county chiefs, few citizens report meeting with health unit management committee members. The index also shows that the largest effect on meetings is due to the information component, but the effect is not significant. While the indices do not suggest an impact on contacts between

Table 8: Treatment-control differences (ANCOVA) - Education

	mean	sc baraza	information	deliberation	level
Number of children in UPS or USE†	1.797	0.149	-0.168	-0.078	0.021
	(1.914)	(0.139)	(0.101)	(0.109)	(0.136)
Distance to public school (km)†	1.420	0.025	-0.047	-0.044	-0.002
	(0.763)	(0.057)	(0.067)	(0.071)	(0.042)
Has complete boundary fence (yes/no)†	0.347	0.064	-0.061	-0.057	-0.008
	(0.476)	(0.048)	(0.046)	(0.049)	(0.045)
Has electricity (yes/no)	0.338	0.165**	-0.040	-0.017	0.035
	(0.473)	(0.049)	(0.042)	(0.049)	(0.038)
Has water facility (yes/no)†	0.703	0.106*	-0.023	0.026	0.073
	(0.457)	(0.041)	(0.048)	(0.050)	(0.050)
Has PTA (yes/no)	0.945	-0.007	-0.029	0.000	0.000
	(0.227)	(0.014)	(0.019)	(0.028)	(0.012)
Has SMC (yes/no)†	0.915	0.008	-0.034	0.002	0.037
	(0.279)	(0.024)	(0.023)	(0.033)	(0.020)
Is informed about SMC (yes/no)†	0.882	0.021	-0.036	-0.042	0.009
	(0.323)	(0.023)	(0.024)	(0.032)	(0.019)
Inspectors visited schools (yes/no)†	0.639	-0.004	-0.075+	-0.035	0.015
	(0.480)	(0.051)	(0.043)	(0.048)	(0.036)
Number of observations	6704	2738	4858	4858	3687

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports the effect (and standard errors below) of the administrative placement of the baraza intervention. **, * and + denotes significance at the 1, 5 and 10 percent levels as assessed using randomization inference. Reported standard errors are clustered at the level of randomization. † indicates that the outcome was in included in the index for the confirmatory analysis. Distance to public school was trimmed at 5 percent and transformed using the inverse hyperbolic sine transformation.

Table 9: Impact of baraza on meetings

	mean	sc baraza	information	deliberation	level
Contact with LC1 chairperson	0.426	0.001	0.030	0.035	-0.034
	(0.495)	(0.025)	(0.035)	(0.048)	(0.024)
Contact with SC chief	0.196	0.031	0.035	0.053	-0.040
	(0.397)	(0.020)	(0.035)	(0.052)	(0.020)
Contact with head teacher/SMC member	0.486	0.038	0.058*	0.048	0.019
	(0.500)	(0.028)	(0.028)	(0.037)	(0.024)
Contact with HUM member	0.155	0.040	0.020	0.061	-0.021
	(0.362)	(0.024)	(0.036)	(0.051)	(0.017)
Contact with Water Committee Member	0.382	-0.016	0.060	0.044	0.012
	(0.486)	(0.040)	(0.034)	(0.051)	(0.037)
Contact Index	0.000	0.037	0.089	0.107	-0.036
	(0.649)	(0.036)	(0.059)	(0.094)	(0.033)
Number of observations	6704	2738	4858	4858	3687

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports the effect (and standard errors below) of the administrative placement of the baraza intervention. **, * and + denotes significance at the 1, 5 and 10 percent levels as assessed using randomization inference. Reported standard errors are clustered at the level of randomization.

citizens and officials, we do find that virtually all coefficients on sub-county level intervention are positive. We also find that a sub-county level barazas are significantly more effective in increasing interactions between citizens and health unit management committee members.

The baraza programme also attempts to increase a sense of community engagement. One way in which citizens can participate is though contributing to common goods such as public infrastructure, education of health services. We differentiate between cash contributions and in-kind contributions. Table 10 shows that about 32 percent of households indicate that they made in-kind contributions to public schools in their community in the last two years. Overall, most in-kind contributions are targeted towards drinking water facilities, and least in-kind contributions were going to dams or irrigation facilities, which is consistent with the difference in public nature of these facilities. Cash contributions are distributed similarly, except for the fact that contributions to bridges and roads generally take the form of labor contributions.

We find that the information component of the sub-county baraza reduces in-kind contributions but increases cash contributions. We further find that cash contributions are significantly higher in areas that were exposed to a district level baraza than in areas that were subjected to a sub-county level baraza. This is driven by significantly higher cash contributions towards schools (and to a lesser extent to drinking water infrastructure). For in-kind contributions, reductions in contributions are especially for schools and for government or community buildings. The increase in cash contributions as a result of the information component of the baraza intervention is especially for drinking water infrastructure.

Finally, we investigate if barazas result in changes in the perception of citizens on a range of problems. Respondents were given a statement, and using a 10 point likert scale, had to indicate how much they disagreed (1) or agreed (10) with the statement. The statements were based on extensive qualitative work where various stakeholders were interviewed and asked about the key problems surrounding public service provision in the different sectors.

Table 11 shows that households tend to agree more that access to drinking water sources is a serious problem as a result of the information component of a sub-county baraza. Households that received a deliberation focused sub-county level baraza are also more likely to agree that drinking water is usually dirty. In the area of public health provision, households that were exposed to a sub-county level information baraza are more likely to agree that access to a health centre or hospital is a serious problem than areas that were exposed to a district level baraza. We further find that households that were exposed to a sub-county level baraza indicate that lack of medicines at health centers or hospitals is less of a problem than in control areas. A direct comparison for this outcome between sub-county level barazas and district level barazas also yields a significant difference. We also ask about perceptions related to friendliness of staff and absenteeism. We find that households that live in areas that received the district level treatment are more inclined to say that absenteeism is a problem, but the difference is not significant.

Table 10: Impact of baraza on contributions

	mean	sc baraza	information	deliberation	level
In-kind contributions to the school?	0.321	0.006	-0.085**	-0.019	-0.059
	(0.467)	(0.032)	(0.032)	(0.035)	(0.031)
In-kind contributions to the health centre?	0.126	0.011	-0.030	-0.031	-0.061+
	(0.332)	(0.023)	(0.025)	(0.021)	(0.023)
In-kind contributions to the road/ bridge?	0.384	0.025	-0.039	-0.011	-0.052
	(0.486)	(0.043)	(0.037)	(0.037)	(0.035)
In-kind contributions to the drinking water facility?	0.452	0.047	-0.010	0.059	-0.004
	(0.498)	(0.046)	(0.042)	(0.038)	(0.043)
In-kind contributions to the dam/irrigation facility?	0.093	0.022	-0.024	-0.028	-0.019
	(0.291)	(0.031)	(0.020)	(0.029)	(0.020)
In-kind contributions to any government structure?	0.233	0.040	-0.073*	0.012	-0.025
	(0.423)	(0.034)	(0.029)	(0.034)	(0.029)
In kind contribution index	0.000	0.063	-0.107+	-0.016	-0.093
	(0.609)	(0.068)	(0.057)	(0.058)	(0.057)
Cash contributions to the school?	0.382	-0.005	0.053	0.021	0.101*
	(0.486)	(0.026)	(0.035)	(0.039)	(0.028)
Cash contributions to the health centre?	0.121	-0.023	0.053+	0.051	-0.014
	(0.326)	(0.024)	(0.035)	(0.040)	(0.017)
Cash contributions to the road/ bridge?	0.097	-0.017	0.001	0.021	0.001
	(0.296)	(0.022)	(0.015)	(0.031)	(0.013)
Cash contributions to the drinking water facility?	0.370	-0.044	0.107*	0.057	0.094
	(0.483)	(0.034)	(0.043)	(0.048)	(0.043)
Cash contributions to the dam/irrigation facility?	0.040	0.001	0.008	0.001	-0.003
	(0.197)	(0.015)	(0.012)	(0.014)	(0.011)
CCash contributions to any government structure?	0.260	0.008	-0.027	0.007	0.029
	(0.439)	(0.030)	(0.026)	(0.039)	(0.030)
Cash contributions index	0.000	-0.033	0.076+	0.063	0.067
	(0.536)	(0.041)	(0.037)	(0.056)	(0.032)
Number of observations	6704	2738	4858	4858	3687

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports the effect (and standard errors below) of the administrative placement of the baraza intervention. **, * and + denotes significance at the 1, 5 and 10 percent levels as assessed using randomization inference. Reported standard errors are clustered at the level of randomization.

We then look at perceptions in the area of education. We see that households are generally most concerned about poor quality learning outcomes, but think absenteeism is less of a problem. For none of the school related perceptions, we find a significant difference between the various groups. We also do not find that the perception of access to roads as a serious problem changes as a result of the barazas.

Respondents seem to perceive agricultural service delivery as the most problematic area: Averages on the likert scales are fairly high when asked if farmers agree extension officers visit rarely, and that there is a lack of transparency in how farmers are selected to receive inputs from government. A perception index that combines all outcome indicates only a significant difference between sub-county level barazas and district level barazas, as district level barazas leads to more disagreement and district level barazas to more agreement.

8 Heterogeneous treatment effects

8.1 Heterogeneity in the timing of the intervention

The slow roll-out of the intervention over an extended period of time also introduces variation in the time that passed between treatment administration and end-line data collection. For instance, the first barazas were held around June 2016 (about one year after the baseline) and so more than 3 years will have passed between treatment administration and end-line data collection. For the most recent barazas, there will only be a few months between treatment administration and end-line data collection. One may argue that sub-counties or districts that were treated early on have been exposed to the program much longer and hence one may expect larger effects on a range of outcomes for these sub-counties or districts than areas that only recently received treatment. At the same time, for some outcomes, effects of the baraza intervention may dissipate (or even reverse) over time as plans are abandoned and promises are not kept.

To capture this, we reran the analysis and added an interaction term between the treatment indicator and an indicator variable that takes the value of one if the baraza that the household was exposed to happened more than 1 and a half years before the endline data collection. Results for the coefficient estimates of the interaction terms are summarized in Figure 6, focusing again on the four hypotheses and the four families of outcomes, and one overall index, similar to the summary in Figure 5[12].

For the agricultural sector, we do not find that the time elapsed between the intervention and endline data collection affects the impact of sub-county level barazas we found in Figure 5. Looking at individual outcomes, we do find that it takes some time for the deliberation component of the baraza to increase the

[12]We focus on summary indices but also discuss results for individual outcomes where relevant. There results are not reported here for brevity, but are available from the authors on request.

Table 11: Impact of baraza on perceptions

	mean	sc baraza	information	deliberation	level
Access to a drinking water source is a serious problem	5.151	0.048	0.606**	0.410	0.143
	(3.264)	(0.265)	(0.223)	(0.227)	(0.160)
Drinking water is usually dirty.	4.428	0.072	0.057	0.442+	0.049
	(3.129)	(0.232)	(0.199)	(0.229)	(0.254)
Access to a government health centre or hospital is a serious problem.	5.819	-0.193	0.365	0.016	0.290
	(3.092)	(0.273)	(0.218)	(0.261)	(0.150)
Government health centres or hospitals do not have relevant medicines.	6.495	-0.412*	-0.027	0.018	0.169
	(3.024)	(0.204)	(0.182)	(0.206)	(0.168)
Staff at government health centres or hospitals are rude to patients.	5.040	-0.048	0.015	0.096	0.053
	(2.913)	(0.224)	(0.155)	(0.205)	(0.165)
Medical staff at government health centres or hospitals are often absent	4.776	0.032	0.081	0.127	0.301
	(2.757)	(0.173)	(0.142)	(0.202)	(0.152)
Access to a government primary school is a serious problem	4.930	0.032	0.046	0.021	0.037
	(2.905)	(0.246)	(0.205)	(0.210)	(0.227)
Teachers in government schools are often absent	4.847	-0.074	0.011	-0.061	0.124
	(2.720)	(0.182)	(0.170)	(0.211)	(0.189)
Children learning outcomes in government schools are poor.	6.360	-0.194	0.166	0.140	-0.246
	(2.918)	(0.180)	(0.155)	(0.187)	(0.154)
Access to all-weather roads is a serious problem	5.157	-0.348	-0.023	-0.180	0.118
	(3.140)	(0.289)	(0.225)	(0.229)	(0.113)
Agricultural inputs supplied by the government are of poor quality.	5.845	0.227+	-0.027	-0.105	0.130
	(2.788)	(0.160)	(0.129)	(0.160)	(0.176)
There is no transparency in inputs distribution.	6.352	-0.351	0.220	0.042	-0.024
	(3.165)	(0.229)	(0.250)	(0.259)	(0.198)
Agricultural extension agents rarely visit.	6.372	-0.189	-0.001	0.103	0.007
	(3.218)	(0.268)	(0.301)	(0.344)	(0.233)
Agricultural extension agents unaware of needs	6.098	-0.010	0.082	0.130	0.458+
	(3.074)	(0.224)	(0.254)	(0.321)	(0.162)
Perceptions Index	0.000	-0.033	0.035	0.026	0.031
	(0.514)	(0.039)	(0.033)	(0.037)	(0.030)
Number of observations	6704	2738	4854	4854	3685

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports the effect (and standard errors below) of the administrative placement of the baraza intervention. **, * and + denotes significance at the 1, 5 and 10 percent levels as assessed using randomization inference. Reported standard errors are clustered at the level of randomization.

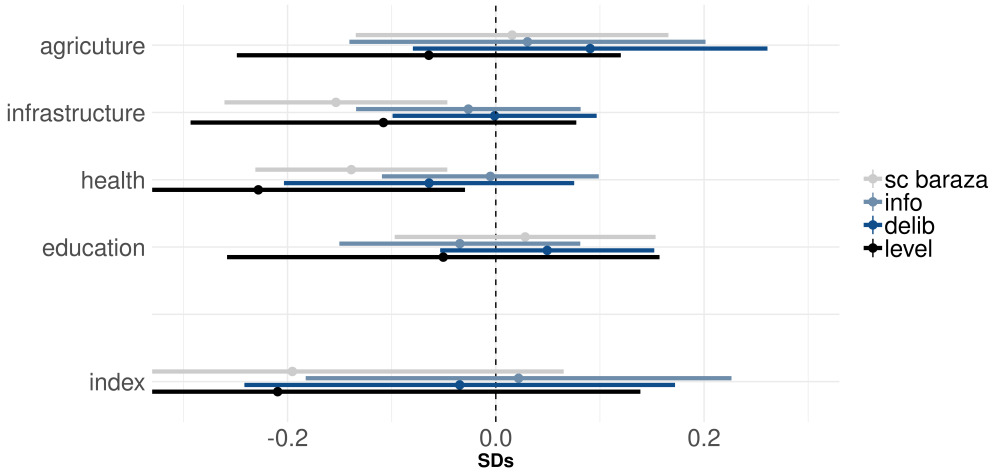

Figure 6: Heterogeneity related to time of intervention

likelihood that households are assisted by cooperatives in marketing. We also find that district level barazas increase the likelihood that households receive NAADS support by almost 5 percent points immediately after the event, but that this positive effect is counteracted by negative effect of almost 10 percentage points after one and a half years.

We find a significant negative interaction effect for the infrastructure index. This is primarily driven by the fact that distance to water source significantly reduces immediately after a sub-county level baraza was held, but increases again after some time, partially reversing the effect. This may indicate that officials start fixing boreholes as a result of complaints raised during the baraza event, but in the longer run the situation worsened as spare parts have been used up.

For outcomes in the health sector, we find negative interaction effects for sub-county level barazas and for district level barazas. We find that the initial strong positive effect of district level barazas on the likelihood of seeking treatment at a government health facility (both for general health care and maternal health cares) disappears in the long run. At the same time, a significant interaction effect of district level barazas on the number of days sick. The negative interaction effect for the sub-county level baraza seems driven by a reduction of the likelihood that households turn to government health facilities for maternal health in the long term. We also find a significant interaction effect for the likelihood of visiting a traditional healer. This is consistent with households being disappointed by the lack of progress made in health centers following the baraza intervention.

For education, we do not find significant interaction effects based for the index. We do find a positive effect of sub-county level barazas on enrollment

in the longer run. This positive long run effect seems due to the information component. We find that that the deliberation component leads to a 12 percentage point increase that the school was visited by an inspector if one allows for sufficient time.

8.2 Heterogeneity related to elite capture

Reinikka and Svensson (2004) find that in the early nineties, schools in Uganda received only about 13 percent of central capitation grants to cover non-wage expenses. While the situation likely improved over the year, elite capture of funds at intermediate levels remains a problem. We look at heterogeneity related baseline levels of elite capture, proxied by the share of the budget that was reportedly received in the year prior to baseline data collection.

Based on data we collected from government officials in all sub-counties, we estimate the average shares of the health, agriculture and education sector budgets that was not received. We find that, overall, only about three quarters of the budget was disbursed. The situation was worse in the agricultural sector, where on average 38 percent of the budget was not received. In the education sector and in the health sector, disbursement rates are the same at 82 percent. For the analysis, we define an indicator variable that takes the value of one if the household lives in a sub-county for which overall disbursement was less than 85 percent and interact this variable with the treatment indicator. Results are reported in Figure 7.

We find that the information component of sub-county level baraza increases overall service delivery in areas characterized by relatively high levels of elite capture. For the agricultural sector, we find that the impact of sub-county level baraza we found in Figure 5 is not subject to heterogeneity. However, we find that the information component of sub-county level barazas becomes significant in areas characterized by lower disbursement rates. This is driven by an increase in the likelihood that households received extension at home.

For infrastructure, we do not find significant interaction effects, although all coefficients are estimated positive. Looking at individual outcomes, we find a significant negative effect from sub-county level barazas on distance to the water source in areas with lower disbursement rates at baseline. We also find a significant negative effect for waiting time as a result of the deliberation component of sub-county level baraza in areas with high elite capture at baseline.

Accounting for elite capture at baseline also leads to a significant impact of the information component in the health sector. This seems to be driven by a substantial reduction in waiting time, but also by an increased demand for public health services.

Finally, for the education sector, we also do not find significant interaction effects. Looking at individual outcomes, we find some indication that district level barazas reduce distance to public schools in areas with lower baseline disbursement rates.

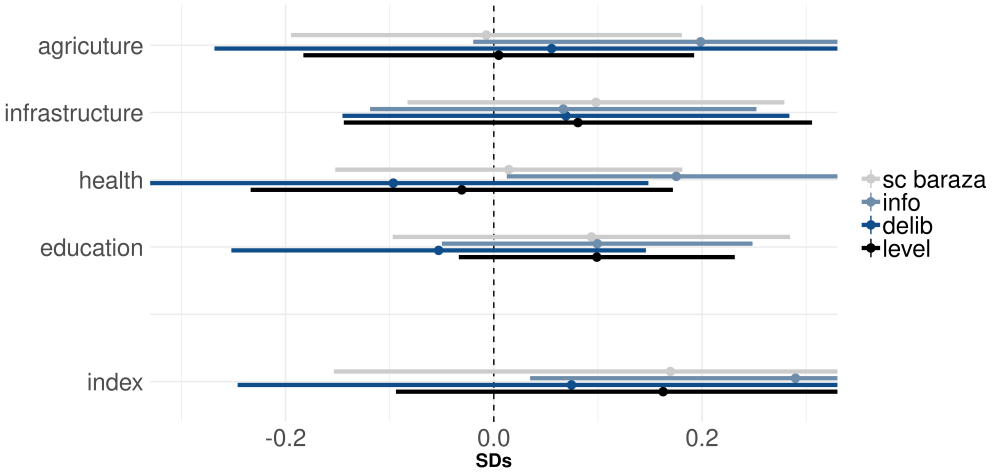

Figure 7: Heterogeneity related to elite capture

8.3 Heterogeneity related to connections to persons in power

Related to the previous, we suspect that connections to people in power can make barazas more effective. There are several reasons why this is the case. For instance, barazas attract a lot of attention, and politicians that hail from a sub-county where a baraza was held may face greater electoral punishment if they are not sensitive to the issues raised. Connections to people in powerful positions may make it easier to direct money from the center to the sub-county to respond to some of the concerns raised during baraza events. Figure 8 thus reports interaction effects of the different interventions with an indicator for sub-counties where officials report if the sub-county has residents that occupy leadership positions such as minister, member of parliament, heads of government agencies, RDCs, and other district level officials.

We do not find any significant interaction effects based on the indices. However, coefficients are generally positive, especially for interactions with the deliberation focused sub-county baraza and district level barazas. Examples of positive interaction effects for individual outcomes include an increased likelihood of visiting extension offices, demonstration sites or model farmers and an increase in the likelihood that schools have a fence after a deliberation focused baraza. We also find a reduction in the use of unprotected water sources after a sub-county level baraza.

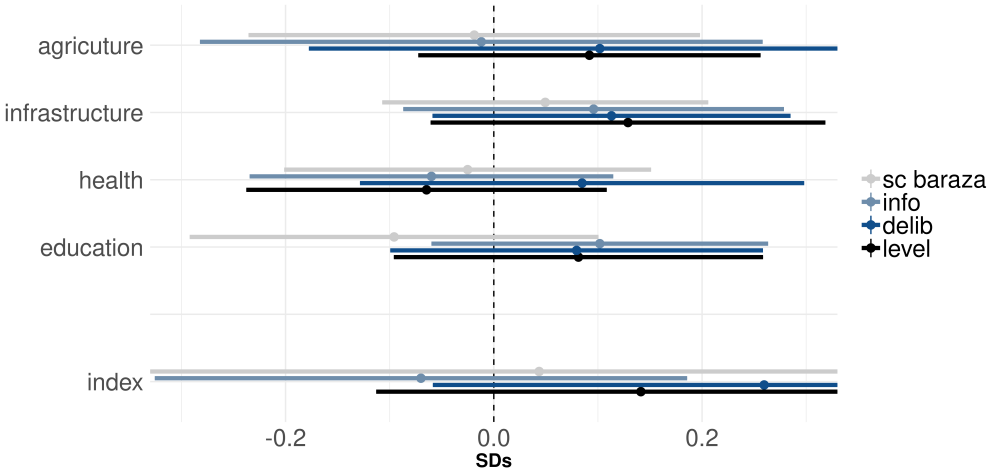

Figure 8: Heterogeneity related to connections to persons in power

8.4 Heterogeneity related to ethnic fractionalization

Building on research showing that ethnic diversity is an important source of variation across communities in the level of public goods provision, Björkman and Svensson (2010) note that free-rider problems inherent to interventions aimed at increasing public action such as the baraza programme may be exacerbated by ethnic fractionalization, as in heterogeneous communities, social norms are likely to be less effective. In addition, they note that in such communities, government officials may also be less sensitive to non-monetary social rewards or sanctions.

In Figure 9, we thus look at interaction effects for the various interventions with a sub-county level indicator for high ethnic fractionalization. To construct this indicator, we follow Björkman and Svensson (2010) and estimate sub-county level measures of ethno-linguistic fractionalization as one minus the sum of squared shares of the different ethno-linguistic groups in the sub-county. We then define sub-counties with high ethnic fractionalization as those with indices in excess of median ethno-linguistic fractionalization.

Contrary to Björkman and Svensson (2010), we find that barazas appear to be most effective in less homogeneous sub-counties. This seems to be particularly the case for district level baraza and combined information and deliberation barazas, and especially for outcomes in infrastructure and the health sector. For the infrastructure sector we find particularly strong interaction effects for distance to water source and distance to all weather roads. Also in the health sector, distance to the nearest government health facility is reduced significantly as a result of both sub-county and district level barazas. In the education sector, we also find that sub-county level barazas reduce distance to

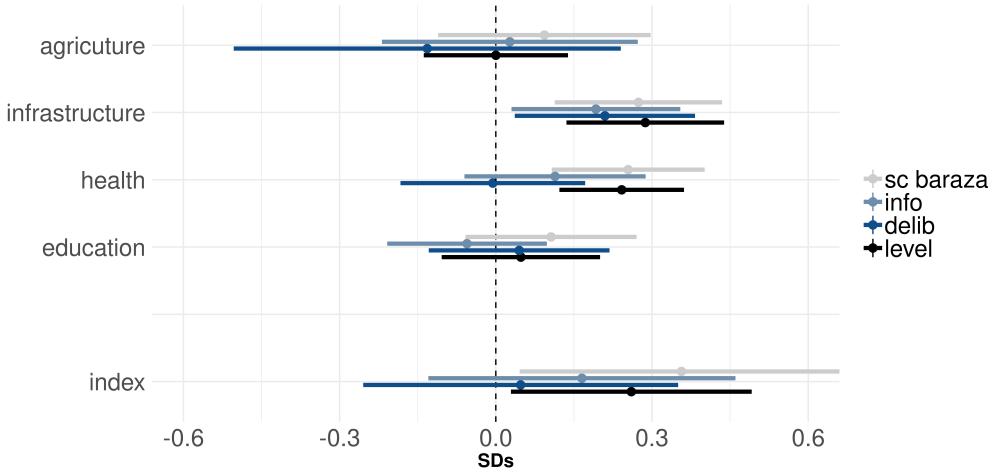

Figure 9: Impact in sub-counties with many ethnic groups

nearest government schools in highly fractionalized communities.

It may be that barazas facilitate the organization of communities in the absence of the social cohesion necessary to engage them in infrastructural works. Consistent with this explanation, we find that distance to the nearest water source, distance to the nearest road, distance to the nearest government health facility and distance to the nearest public school are all significantly higher in control areas with high baseline levels of ethnic fractionalization than in control areas that are more homogeneous.

8.5 Heterogeneity related to remoteness

Heterogeneity related to the timing of the intervention, disbursement rates, connections to persons in power, and ethnic fractionalization all have its source at the sub-county level. Heterogeneity may also be related to household level characteristics. During discussions with stakeholders, it was often argued that barazas may have different effects on households that live close to the sub-county headquarters versus those that live in more remote areas. At baseline, we collected data on the distance between the homestead of the household and the sub-county headquarter. We thus rerun the regressions but now add interaction effects for households that live 4 or more km away from the district headquarter. Results are summarized in Figure 10.

There are some indications that district level barazas are effective for households that live further away from the sub-county. This is particularly the case for outcomes in the agricultural sector. We also find more positive treatment effects in the education sector if we focus on household that live further from the district headquarters.

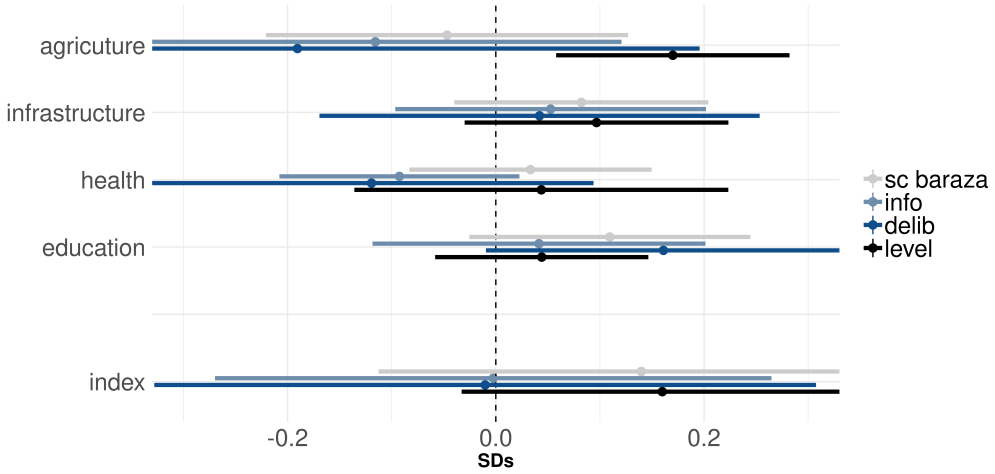

Figure 10: Impact for households living in remote areas

Looking at the agricultural sector more in detail, we find that a district level baraza leads to a significant increase in the likelihood that households are visited by extension officers at home, the likelihood that they visit an extension office, demonstration plot or model farmers, the likelihood of being assisted by NAADS, and the likelihood of being supported by a marketing cooperative.

Remote households also benefit from sub-county level barazas in terms of access to a protected water source. District level barazas reduce waiting time at the source for this sub-group. For access to public health, an overall positive deliberation effect is counter-acted by a negative effect among households living further away from the district headquarter.

9 End-line imbalance

In this final section, we come back to the risk that the partial roll-out may have introduced selection bias. While results in Table 2 and 3 are reassuring, it should be noted that pre-treatment characteristics were collected some time ago and results may be different if more recent data is used and/or if selection happened on characteristics that changed over time. Therefore, we revisit the comparisons between control sub-counties and sub-counties that were allocated to a treatment cell but ended up not being treated, but now use endline data. Specifically for this reason, as mentioned in Section 6.1, instead of simply collecting endline data from the (planned) control sub-counties, we also collected endline data from sub-counties that were supposed to receive a treatment but did not get one (See also footnote 8). Table 12 thus compares end-line outcomes between households that were planned to receive a particular treatment

Table 12: Difference between planned but not treated subcounties and planned controls at endline

	sc baraza	information	deliberation
Agriculture index	0.174**	0.113	0.045
	(0.057)	(0.057)	(0.057)
Infrastructure index	0.026	-0.031	-0.024
	(0.073)	(0.071)	(0.071)
Health index	0.026	-0.028	-0.012
	(0.047)	(0.039)	(0.043)
Education index	0.093	-0.002	0.116
	(0.057)	(0.046)	(0.045)
Public service delivery index	0.161	0.004	0.075
	(0.083)	(0.070)	(0.069)
Number of observations	1637	2356	2808

Note: First column reports differences (and standard errors below) between baseline characteristics of households that were planned to receive a sub-county level combined information and deliberation baraza but did not receive one, and those that were planned not receive any baraza; Column 2 (and standard errors below) reports differences between baseline characteristics of households that were planned to receive a sub-county level information baraza but did not receive one, and those that were planned to not receive a sub-county level information baraza; Column 3 reports differences (and standard errors below) between baseline characteristics of households that were panned to receive a sub-county level deliberation baraza but did not get one, and those that were planned to not receive a sub-county level deliberation baraza; **, * and + denotes significance at the 1, 5 and 10 percent levels. Reported standard errors are clustered at the level of randomization.

but did not end up receiving the treatment to outcomes of households that were assigned to serve as a control for the particular treatment. In the table, we present results for the indices that are also used to summarize impact in Figure 5.

While results need to be interpreted with caution due to small sample size, we do find significant differences between planned but not treated sub-counties and sub-counties that were allocated to the control condition for the agricultural sector. For instance, we find that households that were supposed to receive a sub-county level baraza treatment but did not get one are 10 percentage points more likely to indicate that they visited an extension office, demonstration site or a model farmers. However, as this difference is positive, it could be argued that OPM seemed to have prioritized sub-counties with poorer service delivery in the agricultural sector. As a result, positive results obtained from comparing treated and control groups are likely to underestimate the true impacts of the treatment.

10 Conclusion

To improve governance and public service delivery in Uganda, the Office of the Prime Minister (OPM) organizes community forums—popularly known as barazas—where citizens receive information from government officials, and get the opportunity to directly engage with them. In 2015, we designed a study aimed at evaluating the effectiveness of this policy intervention. The evaluations set out to answer four research questions: (1) what is the impact of the baraza as implemented by the OPM? (2) What is the relative effectiveness of the information component of a baraza? (3) What is the relative effectiveness of the deliberation component of a baraza? And: (4) Should a baraza be organized at the district level or at the sub-county level. Baseline data on more than 12,500 households spread over almost 250 sub-counties in about 40 districts throughout Uganda was collected and OPM started implementing barazas following our experimental design and protocols.

OPM faced various complications that affected the timely roll-out of the barazas, including budgetary constraints and disruptions related to the general elections of 2016. This resulted in the decision to collect end-line information after partial roll-out, more than 4 years after baseline data was collected. Various strategies were followed to test, and reduce the consequences of, potential selection bias introduced by this partial roll out. Early 2020, endline data was collected on 6,700 households.

Overall, following the pre-analysis plan we find that the baraza program did not have a significant impact on public service delivery or its associated outcomes. There are some indications that sub-county level barazas had a positive impact on service delivery in the agricultural sector, and that this impact is significantly larger than the impact associated with district level barazas.

Looking at individual outcomes, we find that households reported increased access to agricultural extension in areas that were exposed to a sub-county level baraza. Furthermore, the deliberation component appeared responsible for an increase in the share of households that reported that they received improved seed from the government extension system. We also found that sub-county level barazas increased the likelihood that farmer cooperatives or groups were formed in the villages in Uganda. Interestingly, it seems that the deliberative component was the main driver behind this result, suggesting deliberation may be important to increase involvement in community affairs. It is also consistent with the previous result, as cooperatives are important vehicles to organize input distribution in Uganda.

For individual outcomes included in the infrastructure index, we find that the time that one must wait at the water source reduced by about 25 percent after a sub-county level baraza took place. The deliberation component seemed to be mostly responsible for this reduction. Furthermore, we again find that the deliberation component seemed to increase community participation, as more households reported to be part of the water user committee. Impact on the health sector was limited. The baraza intervention did seem to be somewhat successful in encouraging sharing of information, as households stated that

VHTs were more likely to organized a public meeting after a sub-county level baraza was organized. Finally, we find that sub-county level baraza seemed to improve school infrastructure.

We also explored if the baraza program affected various aspects that are at the core of community-based monitoring, such as interfacing with politicians and civil servants, perceptions of service quality and prioritization, and contributions to public goods (both cash and in-kind). There is some evidence that the information component of the baraza increased meetings with officials or service providers: even though only a significant effect was found on meetings with the head teacher or members of the school management committee, coefficients on all other indicators were positive and often substantial. We further found that the information component of the sub-county baraza reduced in-kind contributions, but increased cash contributions. This may indicate that barazas increase trust in officials and strengthens the social contract. We further found that cash contributions were higher in areas that were exposed to a district level baraza than in areas that were subjected to a sub-county level baraza. Finally, we found that barazas changed citizens' perception of a range of problems. In general, district level barazas led to more disagreement and district level barazas to more agreement that a range of issues in public service delivery were problematic.

When looking at heterogeneity in treatment effects, we often see that initially positive effects turn negative after some time. This may indicate that investments are made immediately after the baraza events, but resources dry up after some time. It also suggests that over time, households become increasingly frustrated with the lack of progress made against the actions that were promised during barazas. At the same time, some important outcome such as a reduction in morbidity and an increase in school enrollment only materialize after some time. We further find that the information component of sub-county level baraza increases overall service delivery in areas characterized by relatively high levels of elite capture. There is some evidence that sub-counties with connections to people in important government positions benefit more from barazas (especially from barazas that emphasize deliberation), but effects are imprecisely estimated.

We also find that barazas appear to be most effective in less homogeneous sub-counties. In sub-counties characterized by substantial ethnic heterogeneity, health, school and water infrastructure seems to be lacking. This suggests that baraza may be an effective way to mobilize communities that are less likely to invest in public goods due to weak social ties. Finally, we look at heterogeneity related to remoteness and find that particularly district level barazas benefit remote households more than households that live close to the sub-county headquarters.

While we did not find that the baraza impacts public service delivery in general, we did find a variety of interesting effects when we looked at individual outcomes and considered heterogeneity in the treatment effects. The fact that our results are somewhat more encouraging than those found in Raffler, Posner, and Parkerson (2018) may also be related to the fact that our intervention

is organized by the government. Raffler, Posner, and Parkerson (2018) find indications that presence of sub-county officials during the programming boosted the impact of the intervention on treatment quality in health centers. Our analysis thus confirms that top-down monitoring may be more important in changing behavior of civil servants, but that community involvement may be an effective way to pressure officials into increased top-down monitoring.

More generally, our analysis suggests that impact from barazas may be highly localized and context specific. Heterogeneous treatment effects have been documented in similar community-based monitoring interventions, complicating estimation of average treatment effects (eg. Björkman and Svensson, 2010). Barazas are fairly broad interventions that attempt to address a range of issues in a heterogeneous setting, resulting in treatments that are unlikely to be very standardized across treatment units, at least compared to typical treatments in bio-medical sciences. A baraza in one sub-county may evolve in a very different direction than a baraza in another sub-county. Both barazas may be effective in their own right, yet outcomes are averaged over treatment units and impact diluted. This also helps explain diverging findings from a less ambitious qualitative exploration of the likely impact of the baraza programme (Van Campenhout et al., 2018). There, we got the impression that barazas were useful at improving public service delivery across sectors, with stakeholders having no difficulty providing examples of changes they felt were the direct result of the baraza being held: projects that were previously dragging were finished or taken up afresh; sub-standard work was redone; and in some instances, priorities were changed to better align with citizens' needs.

11 Acknowledgment

This project received funding from the International Institute for Impact Evaluations (3ie) under grant number RFQ PW2.18.UG.PG. This research was partly funded by the CGIAR Research Program on Policies, Institutions, and Markets (PIM), led by the International Food Policy Research Institute (IFPRI) and carried out with support from the CGIAR Fund contributors under grant number PIM 203002.002.221. This research was cleared by the Uganda National Council for Science and Technology (UNCST SS 5179), Makerere University School of Social Sciences Research Ethics Committee (MAKSS REC 05.19.291), as well as IPFRI's Institutional Review Board (DSG-19-1053). We would like to thank Alvina Erman, Jennifer Smart, Marc Charles Wanume, Wilberforce Walukano, Fiona Nattembo, Emmanuel Bizimungu, Leocardia Nabwire, Richard Ariong and people at the Office of the Prime Minister, Joseph Muserero in particular.

References

Anderson, M. L. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American statistical Association* 103 (484): 1481–1495.

Arkedis, J., J. Creighton, A. Dixit, A. Fung, S. Kosack, and D. Levy. 2019. "Can Transparency and Accountability Programs Improve Health? Experimental Evidence from Indonesia and Tanzania." .

Banerjee, A., A. Deaton, and E. Duflo. 2004. "Wealth, health, and health services in rural Rajasthan." *American Economic Review* 94 (2): 326–330.

Banerjee, A., R. Hanna, J. Kyle, B. A. Olken, and S. Sumarto. 2018. "Tangible Information and Citizen Empowerment: Identification Cards and Food Subsidy Programs in Indonesia." *Journal of Political Economy* 126 (2): 451–491.

Banerjee, A. V., E. Duflo, and R. Glennerster. 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.

Bertsimas, D., M. Johnson, and N. Kallus. 2015. "The power of optimization over randomization in designing experiments involving small samples." *Operations Research* 63 (4): 868–876.

Björkman, M. and J. Svensson. 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.

———. 2010. "When is community-based monitoring effective? Evidence from a randomized experiment in primary health in Uganda." *Journal of the European Economic Association* 8 (2-3): 571–581.

Björkman Nyqvist, M., D. de Walque, and J. Svensson. 2017. "Experimental evidence on the long-run impact of community-based monitoring." *American Economic Journal: Applied Economics* 9 (1): 33–69.

de Souza Leão, L. and G. Eyal. 2019. "The rise of randomized controlled trials (RCTs) in international development in historical perspective." *Theory and Society* 48 (3): 383–418.

Donato, K. and A. G. Mosqueira. 2016. "Power to the people? A replication study of a community-based monitoring programme in Uganda." *3ie Replication Papers* 11: 92–108.

Francis, P. and R. James. 2003. "Balancing rural poverty reduction and citizen participation: The contradictions of Uganda's decentralization program." *World Development* 31 (2): 325–337.

Fujiwara, T. and L. Wantchekon. 2013. "Can informed public deliberation overcome clientelism? Experimental evidence from Benin." *American Economic Journal: Applied Economics* 5 (4): 241–55.

Gerber, A. and D. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. W. W. Norton.

Goeree, J. K. and L. Yariv. 2011. "An Experimental Study of Collective Deliberation." *Econometrica* 79 (3): 893–921.

Grossman, G. and K. Michelitch. 2018. "Information dissemination, competitive pressure, and politician performance between elections: A field experiment in Uganda." *American Political Science Review* 112 (2): 280–301.

Humphreys, M., R. S. De la Sierra, and P. Van der Windt. 2013. "Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration." *Political Analysis* 21 (1): 1–20.

Kasy, M. 2016. "Why experimenters might not always want to randomize, and what they could do instead." *Political Analysis* 24 (3): 324–338.

Kosec, K. and L. Wantchekon. 2020. "Can information improve rural governance and service delivery?" *World Development* 125: 104376.

Muralidharan, K., M. Romero, and K. Wüthrich. 2019. *Factorial designs, model selection, and (incorrect) inference in randomized experiments*. Tech. rep., National Bureau of Economic Research.

Olken, B. A. 2007. "Monitoring corruption: evidence from a field experiment in Indonesia." *Journal of political Economy* 115 (2): 200–249.

OPM. 2013. *Implementation of the Baraza Initiative. Progress Report*. Tech. rep.

Pandey, P., S. Goyal, and V. Sundararaman. 2009. "Community participation in public schools: impact of information campaigns in three Indian states." *Education Economics* 17 (3): 355–375.

Raffler, P., D. N. Posner, and D. Parkerson. 2018. "The weakness of bottom-up accountability: Experimental evidence from the Ugandan health sector." *Unpublished manuscript* .

Reinikka, R. and J. Svensson. 2004. "Local capture: evidence from a central government transfer program in Uganda." *The quarterly journal of economics* 119 (2): 679–705.

Van Campenhout, B., E. Bizimungu, J. Smart, and N. Kabunga. 2018. "Impact pathways of a participatory local governance initiative in Uganda: a qualitative exploration." *Development in Practice* 28 (8): 1046–1056.