Community advocacy forums and public service delivery - Impact, and the role of information, deliberation, and administrative placement

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

January 7, 2020

Abstract

To improve governance and 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 randomize control trial to assess the impact of the baraza intervention on a range of outcomes related to 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. Furthermore, we compare the effectiveness of barazas organized at the district level to the effectiveness of barazas organized at the sub-county level, as the administrative placement of the barazas is a key determinant of the cost-effectiveness of this policy intervention.

1 Introduction

In 2015, we designed a study aimed at evaluating the effectiveness of community advocacy forums, 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 public service delivery by enhancing public involvement in holding the government accountable for service delivery in relation to the resources spent. The study had several objectives. First, it wanted to establish, in a rigorous way, if the program had an impact

*Development Strategy and Governance Division, International Food Policy Research Institute, 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

on public service delivery. 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), as the level of administrative placement is an important determinant of the cost-effectiveness of the policy intervention. Third, the project also set out to explore pathways through which community advocacy forums may affect outcomes. In particular, using a two-by-two factorial design, it differentiates between the impact of (1) providing citizens with information related to budgeting and planing, and (2) the impact of letting citizens engage with public servants and politicians in a questions-and-answers session. The project was funded by the International Institute for Impact Evaluations (3ie) after a competitive call for proposals, 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 to ensure adherence to the intervention protocols, and the interventions were rolled out by the OPM, our main implementing partner.

While the project was initially assumed to take about 2 years, 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. Four years after the baseline survey, with about 50 percent of the planned barazas implemented, a trade-off needed to be made between waiting for the remaining barazas to be completed or conducting the endline after partial roll-out. It was decided to proceed with endline data collection and employ estimation and data collection strategies to control for potential selection bias that may have been introduced due to the partial roll-out.

The current version of this paper serves as a registered report. In particular, it was prepared as a "mock report", which contains the analysis on simulated endline data for a core set of primary outcomes to test the four primary research questions (impact of sub-county barazas, relative importance of information component, relative importance of deliberative component, and consequences of administrative placement). The outcomes can be categorized into four broad sectors: agriculture, health, education, and infrastructure (including drinking water and roads). With this report, we thus commit to key outcomes in each sector, and combine the outcomes in the four sectors into sector level indices, and in a single index following Anderson (2008) to account for multiple hypothesis testing. Once endline data has been collected, we will simply replace the simulated data with real data, and we will immediately be able to publish the report by running the pre-coded analysis. Pre-registration and mock reports are effective tools against fishing and false-positive science (Humphreys, De la Sierra, and Van der Windt, 2013).[1]

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 look at the impact of a

[1] This document was prepared using Lyx, an open source Latex front-end. All Latex and R code to replicate the analysis is placed under revision control using Git. The R scripts are automatically executed when the Lyx document is compiled (using the R package knitr) and tables are populated. The Git repository can be found at https://github.com/bjvca/baraza/.

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 are still present more than four years after the initial intervention despite minimal follow-up. 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) 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, validates the power of information provision to change the behavior of front-line service providers, but casts doubt on the ability to foster community monitoring or to generate improvements in health outcomes, at least in the short term.

Our study contributes to this literature in various ways. First, this study is one of the 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 (Donato and Mosqueira, 2016). On the one hand, 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 participating officials (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.

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 or international NGO's. It has been argued that successful devolution can only happen in the context of a strong state, able to ensure consistent regulation, and a well-informed public backed up by a participatory political culture (Golooba-Mutebi, 2005). 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 if organized by the government. This is also consistent with suggestive evidence in Raffler, Posner, and Parkerson (2018) 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-allocations 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 (Van Campenhout et al., 2018). Most other studies focus on a single sector; the sector 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)

In this report, we start by providing a brief overview of the government program we study. We motivate the four research questions in a separate section. We then present the design of the original cluster randomized control trial and use baseline information to assess balance. Next, we discuss implementation challenges, and explain strategies we will use to address these. This is followed by as series of updated power calculations. Results are presented based on simulated endline data drawn from the baseline, merely to show what the final analysis would look like. In a final section we outline how endline data will also be used to further assess the likelihood that the implementation challenges introduced selection bias and explore heterogeneity in impact as a result of variation in the timing of the intervention.

2 Background of the baraza impact evaluation

Since Uganda's independence in 1962, Uganda's development efforts have been thwarted by political turmoil and economic mismanagement. 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, for the formulation of services that are more aligned to citizen's needs, and for bringing 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. 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 advocacy forums (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 are 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, it provides the opportunity for citizens to ask questions to their leaders and deliberate among themselves. With barazas, citizens in particular have the opportunity to participate in the policy process by directly engaging with service providers, and to demand accountability for the use of public resources. 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 by now. Barazas were first piloted in the financial year 2009/10.[2] Since then, efforts have been underway to roll out barazas in all subcounties in the country. During the full-scale implementation phase in the financial year 2010/2011, 16 more subcounties in 8 districts had held a baraza meeting. And, by the last quarter of 2011/2012, 267 out of the country's total of 1,340 subcounties, 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 district-level reporting so as to increase participation at a higher level.

A typical baraza is initiated from the center, with the OPM mobilizing district and subcounty officials. These include the Chief Administrative Officer (CAO) as the head of public service delivery 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 (agriculture, education, infrastructure and health). Especially for barazas organized at the subcounty level, the subcounty level equivalents of the CAO (the subcounty chief) and the LC5, the subcounty chairperson (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 venue in which to hold the baraza event. 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

[2]The initial pilot barazas were undertaken in eight lower level local governments (generically referred to as subcounties) of the four districts of Masaka, Bushenyi, Kumi and Nebbi, which are respectively located in the four geographical regions of Uganda: Central region, Western region, Eastern region, and Northern region.

journalists, the RDC seeks accountability and feedback from each head of major sectors. Sector heads are required to present what services were planned to be delivered in the subcounty; 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 reactions and feedback from citizens on whether what has been presented is what 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. In general, a minister of state will also be present at the baraza.

3 Research questions: impact, information provision versus deliberative aspect, and administrative placement

3.1 The impact of (subcounty 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 organized by OPM.

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 classical example of the principle–agent problem. Hence, providing citizens with information about the performance of the agent is an effective way in increasing the quality of public service delivery by allowing citizens to monitor and apply pressure on under-performing 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 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 to 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. Gilens (2001) identifies a significant influence of providing policy facts on the public's political judgment. 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. 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).

3.3 The Deliberation Mechanism

There are various ways in which deliberation increase the quality of public service delivery. Firstly, it has a legitimating 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. Secondly, deliberation can more effectively distill social choice than simple voting and majoritarian rule, in part by building of consensus both among citizens and between public servants and citizens . Thirdly, deliberation has been found to positively impact on the vigor and breadth of subsequent citizen involvement in community affairs (Björkman Nyqvist, de Walque, and Svensson, 2017).

Deliberation also affects 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 fact. 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. A third key question is thus, similar to the previous question, to assess 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 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 had been originally planned to be 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 tradeoff between attempting to achieve breadth of coverage (through the district-level barazas), and attending to depth and quality 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 will be to compare the effectiveness of barazas organized at the higher district level to those organized that the lower subcounty level.

Which is more effective, placement at a higher or lower level, will 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 question, in 2015, 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 subcounties, albeit located throughout all of Uganda's 112 districts across the four regions, had already received a sub-county level baraza intervention. 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 not located in a district in which a district-level baraza had already been implemented. 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.

This study proposed 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 that contain both the information component and the deliberation component $\left(D^{ID}\right)$, while other districts do not receive a baraza at this level $\left(D^{0}\right)$. 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 $\left(S_{ID}^0\right)$. A third quarter will receive a sub-county level baraza that consists largely of officials providing information and limited opportunity for citizens to engage $\left(S_I^0\right)$. A final quarter will receive a sub-county level baraza with a focus on citizens engaging with each other and with officials, without upfront 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 $\left(S_0^{ID}\right)$; these sub-counties do not receive any further baraza at the district level. Within each sub-county, we sample a fixed number of households.

The above design allows us to answer the four research questions from 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 households that were sampled from S_0^0. Second, to inform the government on the consequences of the switch from sub-county to district level barazas, one can compare outcomes of households that were sampled from S_0^{ID} to outcomes of household that were sample from S_{ID}^0. Third, 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, because of the factorial design, much more information can be used to test the two last hypotheses than for the two first hypotheses.

To determine the number of districts, sub-counties and households to include in the study, the original research proposal contained 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

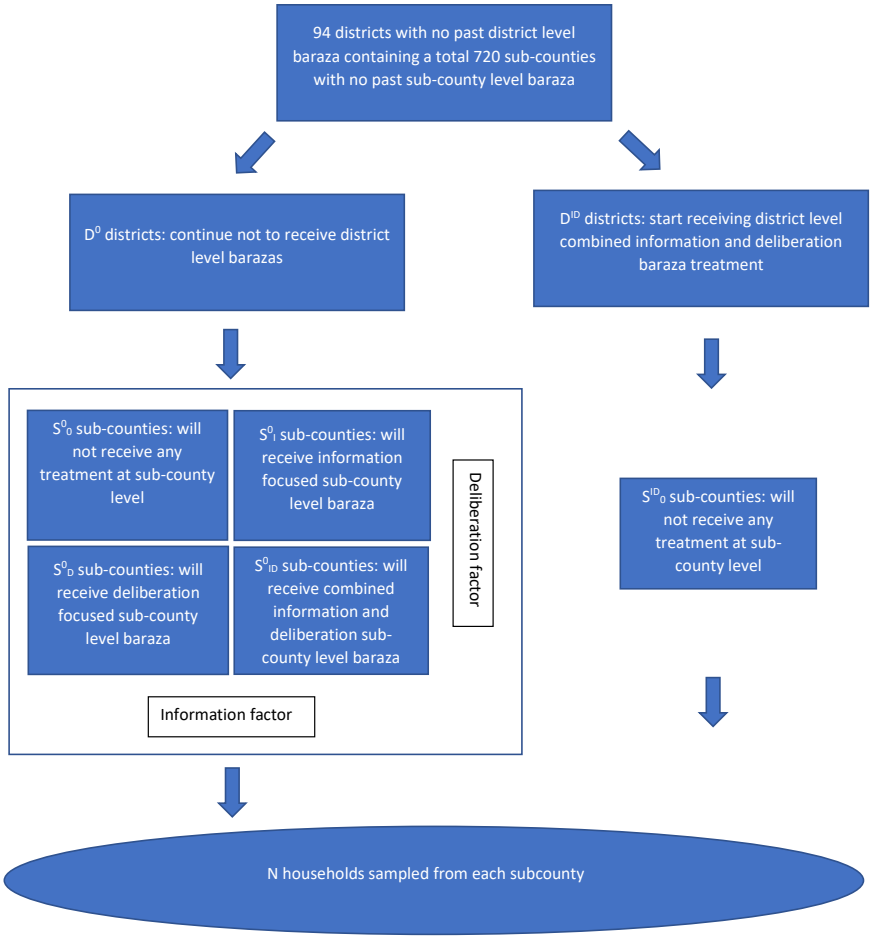


Figure 1: Experimental Design

errors of the outcomes and interclass 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 distributed over 230 subcounties in 40 districts throughout Uganda, on which baseline data was collected[3]. More details on the power calculations can be found in the original proposal, which is available as an online appendix. In this document, we will run a series of updated power calculations to account for the implementation challenges (see Section 5)

In Table 1, we test for balance between the treatment groups at baseline following the initial design of the experiment. During baseline, information on 12,545 households was collected. 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. In the second column, we report differences between baseline characteristics of households that will receive a sub-county level combined information and deliberation baraza, and those that will not be exposed to any baraza. We can not reject the null that households in these two groups are similar for any of the characteristics in Table 1. When comparing households that were exposed to a sub-county level information baraza to households that did 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 at the 5 percent level. 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, but the difference is only significant at the 10 percent level.

In the fourth column of Table 1, we report differences between households that were exposed to a sub-county deliberation baraza and households that were not. For this treatment, we can not reject balance on any of the variables. In the last column, we report differences in outcomes between households that were exposed to a district level baraza and households that were exposed to a sub-county level baraza that combined both information and deliberation components. We see that household heads in the first group are slightly older than in the latter group. Furthermore, the share of households that report that there is a Village Health Team in their village is also slightly higher in the treatment group. In both cases, the difference are significant at a 10 percent level. Overall, out of 40 comparisons, we find that one difference is significant a the 5 percent level and three are significant at the 10 percent level, which is what one would expect to find due to chance alone. As such, we conclude that the initial randomization was successful.

[3]We added an additional 3 subcounties 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	-0.284	0.304*	-0.003	0.246
	(2.825)	(0.190)	(0.133)	(0.125)	(0.248)
Age of the household head (years)	46.500	-0.474	0.464	0.725	1.432+
	(14.612)	(0.901)	(0.594)	(0.714)	(0.801)
Head of household is woman (1=yes)	0.191	0.023	-0.014	0.004	-0.013
	(0.393)	(0.021)	(0.013)	(0.015)	(0.016)
Head finished primary education (1=yes)	0.213	0.016	-0.020	-0.003	-0.026
	(0.410)	(0.027)	(0.020)	(0.020)	(0.027)
Thatched grass roof (1=yes)	0.298	0.021	0.009	-0.032	0.011
	(0.457)	(0.034)	(0.025)	(0.023)	(0.036)
Traditional mud wall (1=yes)	0.424	0.007	-0.025	0.038	-0.034
	(0.494)	(0.058)	(0.040)	(0.039)	(0.104)
Distance to nearest all weather road (km)	0.906	-0.088	0.106	0.147	-0.192
	(0.915)	(0.145)	(0.095)	(0.092)	(0.138)
Access to extension (1=yes)	0.108	-0.008	0.004	0.007	0.009
	(0.310)	(0.018)	(0.012)	(0.014)	(0.017)
Village Health Team in village (1=yes)	0.854	-0.031	0.006	0.025	0.070+
	(0.353)	(0.039)	(0.026)	(0.026)	(0.036)
Number of children in public schools	2.478	-0.161	0.165+	0.038	0.139
	(2.074)	(0.128)	(0.091)	(0.089)	(0.155)
Number of observations	12545	10241	10241	10241	4949

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the subcounty 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.

5 Implementation challenges and subsequent design adaptations

One of the main challenges was a slow roll-out of the Baraza intervention by the implementing partner. At the start of 2018, and almost two and a half years after baseline data was collected, only about 25 percent of the planned interventions had happened, and we needed to balance the costs and benefits of waiting until OPM finished all barazas or collecting baseline information after incomplete roll-out. At that time, we developed various scenarios, each with an adapted research design. After an additional six months, with still only 56 out of the 155 Barazas implemented, it appeared that the best scenario would be one whereby endline data would be collected before all subcounties were treated.

However, end-line data collection after partial roll-out may introduce selection bias. It may be that, from the randomly assigned sub-counties, particular sub-counties were selected to be treated first and others postponed. For instance, the implementing partner may have started with sub-counties that are close to the capital due to logistical reasons. Below are three strategies we will follow to address the issue of potential selection bias.

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

First, we can investigate if selection bias was introduced 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.[4] 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.

Table 2 presents the original balance table (Table 1), but after dropping sub-counties that were treated. Thus, instead of comparing pre-treatment characteristics between treatment sub-counties and control sub-counties, the table compares sub-counties that were allocated to a particular treatment (but did not end up receiving the treatment) to the control for that particular treatment. The table seems to suggest that the roll-out did not introduce imbalance, at least as judged by the pre-treatment characteristics that were in the original balance table. We find that, out of 30 comparisons, we reject the null hypothesis of no difference at a 5 percent significance level twice and at 10 percent once. Also

[4] All district level barazas were implemented, so we only focus on sub-counties here and in the following sections.

here, this would be expected by pure chance alone, and so we can concluded that the partial roll-out did not seem to have introduced selection bias.

Note also that this can again be tested at end-line, where we would expect to find similar post-treatment outcomes in the subset of sub-counties where a baraza was planned and not implemented, as well as in the subset of sub-counties where a Baraza was not planned (control Barazas). We implement this test in Section 8.2 below.

5.2 Selection of control subcounties to be included in end-line survey

Second, as only part of the intervention was implemented, it will 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). Indeed, statistical power is likely to be highest for an equal number of treated and control sub-counties, and while adding more control sub-counties will increase precision, the gains are unlikely to outweigh the cost. 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 data collection? One reasonable suggestion would be to pick them 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 example, it may be that the implementer prioritized sub-counties that were closer to the capital. Randomly selecting control sub-counties may mean that sub-counties closer to the capital are relatively under-represented and sub-counties that are further away may be 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 affect his or her decision (Kasy, 2016; Bertsimas, Johnson, and Kallus, 2015). For instance, on the basis of the GPS coordinates of a treated sub-county, a control sub-county that is relatively close to the treated sub-county can be selected from the different candidate control sub-counties.

We decided to use 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

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

	mean	sc baraza	information	deliberation
Household size	6.324	-0.315	0.388*	0.023
	(2.825)	(0.215)	(0.170)	(0.140)
Age of the household head (years)	46.500	-0.658	0.691	0.567
	(14.612)	(0.933)	(0.664)	(0.808)
Head of household is woman (1=yes)	0.191	0.020	-0.019	-0.003
	(0.393)	(0.023)	(0.016)	(0.017)
Head finished primary education (1=yes)	0.213	0.016	-0.007	-0.003
	(0.410)	(0.028)	(0.027)	(0.022)
Thatched grass roof (1=yes)	0.298	0.017	0.000	-0.036
	(0.457)	(0.036)	(0.024)	(0.027)
Traditional mud wall (1=yes)	0.424	0.001	-0.058	0.044
	(0.494)	(0.064)	(0.047)	(0.044)
Distance to nearest all weather road (km)	0.906	0.037	0.009	0.188+
	(0.915)	(0.165)	(0.100)	(0.110)
Access to extension (1=yes)	0.108	-0.002	0.008	0.007
	(0.310)	(0.019)	(0.016)	(0.015)
Village Health Team in village (1=yes)	0.854	-0.032	-0.010	-0.014
	(0.353)	(0.043)	(0.028)	(0.028)
Number of children in public schools	2.478	-0.176	0.249*	0.076
	(2.074)	(0.143)	(0.115)	(0.100)
Number of observations	12545	9241	7792	8341

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the subcounty 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. **, * and + denotes significance at the 1, 5 and 10 percent levels.

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

Figure 2: Factorial design

of parliament coming from the sub-county). These characteristics are used in a probit regression to predict the likelihood that a sub-county was treated. For each treated sub-county, we then match a potential control sub-county with a likelihood of being treated that is similar to that of the treated sub-county.

Figure 2 summarizes the factorial design that underlies the assessment of the relative effectiveness of the information and deliberation components of subcounty level barazas. As already note above, one of the main advantages of factorial designs (as opposed to parallel designs) is the fact that, to test main effects, all observations can be used. For instance, to test the impact of an information Baraza, we can compare outcomes of households in sub-counties that received the information treatment (either only the information treatment or the information + deliberation treatment) to outcomes of households that did not receive the information treatment (either because they received no treatment at all or because they only got the deliberation treatment). If the intervention had been implemented as planned, we would thus have had 104 information sub-counties that could be compared to 102 control sub-counties (and as 50 households were interviewed in each sub-county, we would have 5,200 treated households and 5,100 control households).

However, the incomplete rollout resulted in the fact that only 67 of a total of 155 sub-counties that would have received any treatment were actually treated. Referring to Figure 2, we see that to test the impact of the information Baraza, 49 sub-counties that were treated can be used. This means that a total of 157 sub-counties that did not receive the information treatment can be used as control sub-counties. However, optimal power is obtained in designs where the number of treated units is about equal to the number of control units, so from a cost-efficiency perspective; we will collect information on 49 sub-counties. As we want to formally test if the partial roll-out introduced selection bias by comparing planned control sub-counties to sub-counties that were not treated using endline data (Section 8.2), we will make sure we select half of these from the first column in table 2, and half from the second column. To test the impact of the deliberation treatment, we need 38 control households. Also here, we will make sure half are from the planned controls (first row in table 2) and half from sub-counties that were supposed to be treated, but were not (second row). Finally, as we also plan to directly test for the effect of a combined information+deliberation treatment, we will need at least 20 pure control sub-

counties. Also here, we will make sure half are selected from the upper left cell in table 2 and half from the sub-counties that were assigned to the treatment in the lower right cell of table 2 but did not get the treatment. Note that often, the same sub-county can be used to test different hypotheses. For instance, the 10 sub-counties in upper left cell needed to test if the deliberation intervention was effective can be taken from the 14 sub-counties that are needed in that cell to test the impact of the information treatment. We thus simply take the higher number in each cell, which is 14 sub-counties. To allow for attrition, we will select 16 control sub-counties in each treatment cell.

In practice, we started by matching 10 untreated sub-counties from from the S_0^0 group to the treated sub-counties in the S_{ID}^0 group. We then match a further 10 sub-counties from the S_{ID}^0 group that ended up not being treated to the treated sub-counties in the S_{ID}^0 group. Next, we look at the information treatment. In this treatment, 49 sunbcounties have been treated, either as information alone or as part of the combined information and deliberation treatment. This means we also need 49 controls. We already have selected 20 pure controls in the previous step which we can used. Furthermore, 18 pure deliberation treatments can be used as controls for the information treatment as well. This means we need an additional 11 controls. As we want to investigate balance between control and planned but not treated controls, we select these 11 controls from the subcounties that were planned to receive the information treatment S_I^0 but ended up not receiving the treatment.

Finally, we look at the deliberation treatment. In this treatment, 38 subcounties have been treated, either as deliberation alone or as part of the combined information and deliberation treatment, so we also need 38 controls. We already have the 20 pure controls and an additional 11 controls from the previous. So we need an additional 7 controls. As we want to investigate balance between control and planned but not treated controls, we select these 7 controls from the sub-counties that were planned to receive the deliberation treatment S_D^0 but ended up not receiving the treatment.

In Table 3, we look at baseline balance for this updated sample (Table 3). Interestingly, the imbalance that was found in Table 1 for the information treatment on household size and the number of children in school has disappear: apparently, the sub-counties responsible for this imbalance ended up not being treated (Table 2). Also here, across 40 comparisons, we would expect to find 2 significant coefficients at 5 percent level and 4 at the 10 percent level. Hence, 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.

5.3 Analysis: matched difference-in-difference

A third way in which we may want to account for selection bias introduced by the partial roll-out is by adapting the way in which impact is estimated. While we will rely on analysis of covariance (ANCOVA) as the main specification to assess impact, we will also report results based on matched difference-in-

Table 3: Orthogonality tests for final sample

	mean	sc baraza	information	deliberation	level
Household size	6.407	0.019	0.065	-0.300^{+}	0.194
	(2.856)	(0.243)	(0.153)	(0.166)	(0.212)
Age of the household head (years)	46.975	0.275	-0.203	0.620	0.545
	(14.545)	(1.398)	(0.732)	(1.042)	(0.997)
Head of household is woman (1=yes)	0.191	0.009	-0.006	0.022	-0.008
	(0.393)	(0.031)	(0.018)	(0.024)	(0.017)
Head finished primary education (1=yes)	0.208	0.014	-0.016	0.015	-0.049^{+}
	(0.406)	(0.042)	(0.025)	(0.035)	(0.029)
Thatched grass roof (1=yes)	0.260	-0.024	0.044	-0.006	0.039
	(0.439)	(0.044)	(0.030)	(0.022)	(0.050)
Traditional mud wall (1=yes)	0.442	-0.027	0.031	0.063	-0.144
	(0.497)	(0.085)	(0.053)	(0.058)	(0.098)
Distance to nearest all weather road (km)	0.905	-0.205	0.027	-0.103	0.080
	(0.910)	(0.198)	(0.140)	(0.135)	(0.127)
Access to extension (1=yes)	0.105	-0.007	0.000	0.013	0.009
	(0.307)	(0.025)	(0.012)	(0.020)	(0.021)
Village Health Team in village (1=yes)	0.863	-0.107^{+}	0.020	0.091*	0.111*
	(0.344)	(0.059)	(0.036)	(0.039)	(0.049)
Number of children in public schools	2.506	0.091	0.001	-0.188^{+}	0.184
	(2.073)	(0.163)	(0.097)	(0.110)	(0.130)
Number of observations	12545	5348	5348	5348	3000

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the subcounty 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.

difference. The decision on which specification is to be preferred in case of diverging results will also depend on the conclusions drawn from the comparison between planned but not treated sub-counties and planned control sub-counties at endline (see section 8.2).

For the matched difference-in-difference estimator, we use Mahalanobis distance with coarsened exact matching, an extremely powerful method of matching (Iacus, King, and Porro, 2012). We match on household size, sex of the household head, age of the household head, whether the household head finished secondary education, the logarithm of farm size, housing conditions (iron roof and improved wall), phone ownership, latitude, and longitude.[5] Endline data is then merged to the matched dataset, and standard difference-in-difference models are then estimated.

6 Updated power calculations

The original power calculations assumed full roll-out of the intervention. The partial roll-out necessitates an update to these power calculations to obtain a new set of minimal detectable effects (MDEs) associated with the sample that will be collected. 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).

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 report that they were visited by an extension officer in the last year. The gray solid line shows the power curve associated wiht the deliberation treatment, comparing the 1,900 households that received the information treatment to the 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 the 2,450 households that live in subcounties that received an information baraza to the 2,900 households that did not receive a subcounty information baraza. The dark blue dashed line compares effectiveness of barazas conducted at different levels, with the MDE defined as the difference in outcome between 1,000 households that received the combined information and deliberation subcounty level baraza and 2,000 households that were exposed to a district level baraza. Finally, we also investigate power for the comparison between pure control barazas and the sub-county level baraza (black dotted line). Here we compare 1,000 households

[5] For the coarsened exact matching, custom cut points were defined to construct 3 age categories, six farm size categories and a five-by-five grid based on coordinates. For the comparison between sub-county level barazas and district level barazas, we did not match on latitude and longitude, as this resulted in too many observations that could not be matched.

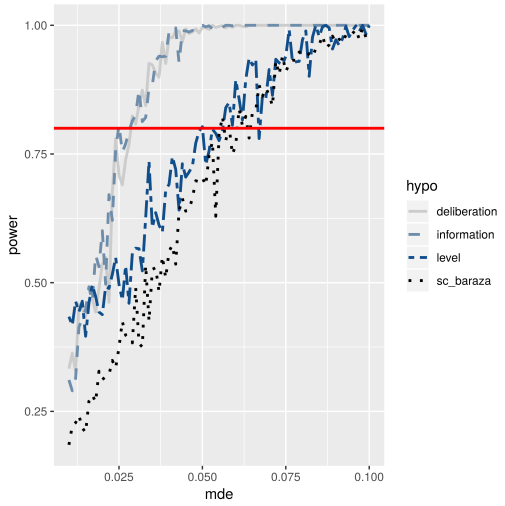


Figure 3: Power curves for access to extension

that received the combined information and deliberation subcounty level baraza to the 2,000 households that did not receive any baraza. MDEs are estimated using a simple ANCOVA model that controls for the outcome at baseline.

Not surprisingly, we have most power for testing the information treatment. We see that the power curve hits the 80 percent threshold a first time at an MDE of about 2.5. The deliberation experiment is similarly powered, and at 80 percent we can expect to identify effects of 3 percent of more. Due to the smaller sample size, directly comparing sub-county level barazas to district level barazas seems harder. Here, the difference needs to be at least 5 percent[6]. For similar reasons, it is much harder to compare pure control sub-counties to sub-county barazas that received the interacted treatment.

In figure 4, we plot MDEs for an infrastructure related outcome: distance (in km) to the primary water source during the dry season. We find that for the information treatment and the deliberation treatment, we can detect an 4 percent difference at the standard 80 percent power level. As the average household lives about 900 meters from the primary water source, this means we can identify effects in excess of 36 meters. Also here, the MDE is higher when directly comparing the effect of district level barazas to subcounty level barazas or when we assess the impact of the sub-county level barazas as implemented

[6]Unless effects from district level barazas and sub-county level barazas go in opposite directions, it seems unlikely that we will be able to detect a difference in a direct comparison. Even though a direct comparison is most interesting and most relevant from a policy perspective, we have more power for other comparisons. For instance, we can also compare outcomes of the 2,000 households that were exposed to a district level baraza to the 2,000 households that did not receive a baraza. Alternatively, we can compare outcomes of the 2,000 households that were exposed to a district level baraza to the 3,350 households that received any type of sub-county level baraza (information only, deliberation only, or the combined).

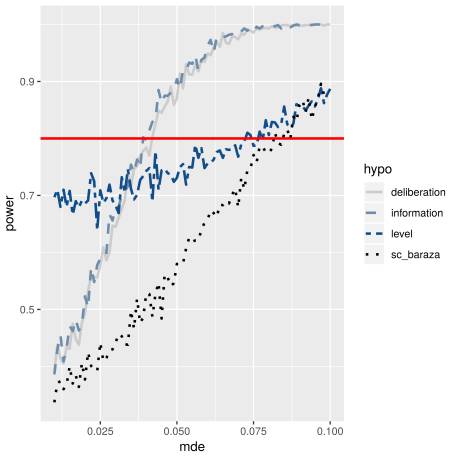


Figure 4: Power curves for distance to water source

by the OPM. Then, MDEs correspond to about 70-90 meters for the average household in our sample. In Appendix 11.1, we run similar analysis for all the variables that we will use to judge impact of the baraza intervention.

7 Results

In this section, we provide results for the four main hypotheses outlined in Section 3. For now, the tables and graphs are generated by running the code on a simulated endline data set[7]. Once endline data is collected, we simply replace the simulated endline dataset with the actual endline. Preparing and pre-registering such a "mock report" reduces the likelihood that results are driven by specification search.

This section will provide clear definitions of what variables will be used to assess impact. For continuous variables, 5 percent trimmed values will be use (2.5 percent trimming at each side of the distribution). Inverse hyperbolic sine (IHS) transformations will be used if skewness exceeds 1.96. Trimming will always be done on end results. For instance, if the outcome is yield at the plot level, then production will first be divided by plot area, after which the IHS transformation is done, and the end result is trimmed. Outcomes for which 95 percent of observations have the same value within the relevant sample will be omitted from the analysis to limit noise caused by variables with minimal variation.

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

[7] The simulated endline was just a random draw from the baseline data

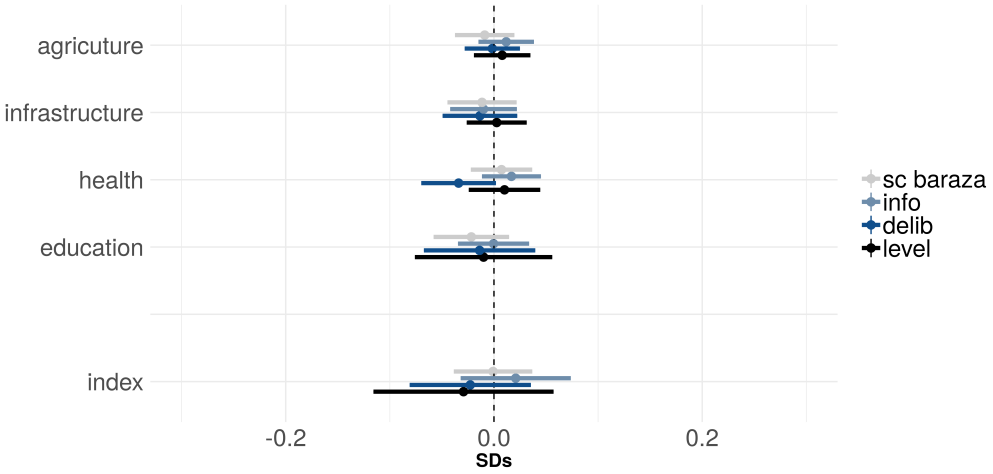


Figure 5: Summary of baraza impact

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 there results from the baraza impact evaluation. It shows the impact of the four main hypotheses–the impact of the subcounty baraza, the relative effectiveness of the information component, the relative effectiveness of the deliberation component, and a comparison between sub-county and 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, which are discussed in detail below. We also combine the four indices into one overall index that assesses the impact on public service delivery in general.

7.1 Agriculture:

We now zoom in on the outcomes that will be considered in assessing the effectiveness of baraza to change service delivery in agriculture. A first outcome looks at access to extansion at home. In particular, 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 (the variable named "baraza.B2" in the endline questionnaire). We find that access to extension is low, with only about 11 percent of households reporting that they received such a visit. However, while extension officers may not visit households, households may still have access to information if they are able to visit extension offices,

demonstration sites or model (baraza.B3 or baraza.B3.3). This percentage is even lower at baseline (8 percent).

Extension services used to be provided under the National Agricultural Advisory Services (NAADS). These services encouraged the formation of farmer associations and groups, which could then request training and inputs from NAADS. Therefore, the presence of NAADS supported farmer groups is also a useful indicator of agricultural service delivery in Uganda (baraza.B4.1). Recently, NAADS has been taken over by the army and is now known as Operation Wealth Creation (OWC), which focusses more on input delivery and less on agronomic advice. We find that at baseline, about 18 percent of households report that there is a NAADS/OWC supported farmer group in their village.

Further down impact pathway is actual change in the use of modern inputs by farmers in areas where agricultural related services are improved as a consequence of the baraza interventions. For instance, we also estimate the proportion of households in our sample that report to have used inorganic fertilizers (DAP, Urea, NPK, Foliar, TSP, SSP, MOP) or improved seed in the last 12 months (baraza.B1 or baraza.B1.5). In Uganda, the use of modern inputs is very low; at baseline, we find that only 34 percent of households used inorganic fertilizers of improved propagation material in the previous year.

The ministry of agriculture does not only provide services aimed at increasing production and productivity. Connecting farmers to markets is also an important strategy outlined in the Agriculture Sector Strategic Plan (ASSP). Therefore, we also include two outcomes that look at public services related to crop marketing. First, we estimate the proportion of households in our sample that report they received help in marketing their produce from the village procurement committee/village farmers forum in the last 12 months. (baraza.B5.2) Second, we ask a similar question to assess the proportion of households in our sample that report they received help in marketing their produce from a cooperative or association in the last 12 months. (baraza.B5.3). At baseline, these percentages are, respectively, 44 and 27.

Results of the ANCOVA models for each of these outcomes are reported in Table 4. The first column reports baseline averages, with standard deviation in brackets below. In the second column, we report difference in outcome between households that received a typical 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). 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 deliberation baraza; top and bottom left of Figure 2). 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

Table 4: Treatment-control differences (ANCOVA) - Agriculture

	mean	sc baraza	information	deliberation	level
Was visited by extension officer at home (yes/no)	0.112	0.014	0.000	0.005	-0.007
	(0.316)	(0.017)	(0.010)	(0.008)	(0.010)
Visited training or demonstration site (yes/no)	0.080	0.002	-0.008	0.000	-0.008
	(0.271)	(0.014)	(0.010)	(0.013)	(0.009)
NAADS or OWC in village (yes/no)	0.179	-0.029*	0.019	-0.006	0.012
	(0.384)	(0.012)	(0.012)	(0.016)	(0.013)
Uses modern inputs (improved seed or fertilizer) (yes/no)	0.339	-0.004	0.002	-0.003	0.001
	(0.473)	(0.013)	(0.013)	(0.015)	(0.016)
Support in marketing from village procurement committe (yes/no)	0.439	-0.027+	0.004	0.008	0.015
	(0.496)	(0.015)	(0.016)	(0.018)	(0.022)
Support in marketing from cooperative (yes/no)	0.269	0.010	0.016	-0.008	0.017
	(0.444)	(0.014)	(0.014)	(0.015)	(0.015)
Agriculture Index	0.000	-0.009	0.012	-0.002	0.008
	(0.409)	(0.014)	(0.013)	(0.013)	(0.013)
Number of observations	12545	5321	5321	5321	2988

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the subcounty 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.

that were not exposed to a deliberation baraza (either pure control or information only baraza; top left and right of Figure 2). Finally, in the fourth column, we directly compare 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) to households that live in sub-counties that were exposed to a district level baraza.

7.1.1 Infrastructure

This study focuses primarily on water and road infrastructure. A first outcome we consider is whether the household uses unprotected water source during dry season (yes/no). This is measured as the share of households that report that the main source of drinking water during the dry season is rain water, surface water, water obtained from a tube well or borehole, an unprotected dug well or and unprotected spring. (baraza.C1). We find that overall, 70 percent of households rely on unsafe water during the dry season.

Next, we look at the distance to the primary water source (baraza.C1.2) and waiting time at the water source (baraza.C1.3), both during the dry season. The first is measured in km, and we see that the average villager needs to walk about 0.91 km and needs to wait on average 36.24 minutes according to data collected at baseline. We also ask if there is a water user committee in the village (baraza.C2.3) and find that at baseline,about 53 percent of households answer affirmative to this question.

We include one question related to road infrastructure. We ask how far the household is located from the nearest all weather road (in km; baraza.A6). We find that at baseline, a household lives on average 1.76 km from a road. Table 5 provides details of the impact of the baraza intervention on infrastructure. The table is organized similar to Table 4.

7.1.2 Health

The first two outcomes we consider attempt to assess changes in access or use of public health facilities. A first indicator measures the use of public health facilities 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 your household had fever (baraza.D2). A similar indicator attempts to assess the use of the public health system 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 your household was to give birth (baraza.D2.4).

Next, we ask if a Village Health Team is present in the village (baraza.D3). VHTs are very important in front-line health care in Uganda. They also have prominent roles in government health interventions, such as immunization campaigns or the distribution of bed nets. We find that at baseline 75 percent of households report that a VHT is present in their village. We also consider distance to the nearest government health facility, measured in km (baraza.D4.2).

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

	mean	sc baraza	information	deliberation	level
Household uses unprotected water source during dry season (yes/no)	0.696	0.013	0.020	-0.012	0.013
	(0.460)	(0.014)	(0.018)	(0.019)	(0.015)
Distance to water source	0.708	-0.010	-0.010	-0.033	-0.006
	(0.528)	(0.016)	(0.019)	(0.022)	(0.017)
Average waiting time at source (min)	2.712	0.007	0.132*	0.221**	-0.061
	(2.169)	(0.015)	(0.066)	(0.081)	(0.097)
Is there a Water User Committee in this village? (yes/no)	0.525	-0.004	0.014	-0.006	0.031*
	(0.499)	(0.014)	(0.017)	(0.018)	(0.015)
Distance to nearest all weather road (km)	0.896	0.007	0.002	-0.050	-0.014
	(0.912)	(0.013)	(0.026)	(0.037)	(0.033)
Infrastructure Index	0.000	-0.011	-0.010	-0.013	0.003
	(0.449)	(0.017)	(0.016)	(0.018)	(0.013)
Number of observations	12545	4763	4763	4763	2716

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the subcounty 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.

At baseline, average distance to the nearest government health facility is 4.66 km. This outcome is included as a negative outcome in the indices, as we expect barazas to reduce the distance.

We then ask whether any household members were unable to work or go to school due to an illness in the past one year (baraza.D1). We then ask how long did you have to wait before being attended (in min) (baraza.D4.6). At baseline, we find that the average patient has to wait about 85 minutes before being examined. Finally, we ask if a traditional health practitioner was visited in the last year (baraza.D6). We find that, at least at baseline, few households use traditional health practitioners (about 9 percent of households). Results for the four comparisons corresponding to the four primary research questions are reported in Table 6.

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.

7.1.3 Education

The problem with education as an outcome is that 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. In fact, because of this, it was decided not to include the last two outcomes in this family.

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 (UPE; baraza.E1.2) or Universal Secondary Education (USE; baraza.E2.1)). Baseline data suggests that the average household in our sample had on average 249.81 children in government schools. 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; baraza.E1.2 and baraza.E2.2). We find that on average, households live about 3 km from a government operated school.

We also look at school infrastructure. First, we ask households if the the primary or secondary school attended by any of their children has a complete boundary fence (baraza.E1.4 and baraza.E1.4). At baseline, it was reported that only about 31 percent of schools have such a fence. We also ask if there is a water source available in the school (baraza.E1.6 and baraza.E2.6). At baseline, this seemed to be the case in the majority of the schools, as about

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

	mean	sc baraza	information	deliberation	level
Seek treatment for fever in public health facility (1=yes)	0.736	0.018	0.017	-0.002	0.015
	(0.441)	(0.012)	(0.014)	(0.018)	(0.015)
Go to public health facility to give birth (1=yes)	0.840	0.006	-0.010	-0.010	0.013
	(0.366)	(0.014)	(0.012)	(0.016)	(0.013)
Is there a VHT in village? (1=yes)	0.752	0.004	-0.011	-0.028	0.021
	(0.432)	(0.016)	(0.016)	(0.019)	(0.015)
Distance to nearest govt health facility (km)	1.998	0.018	0.015	0.069*	-0.028
	(0.745)	(0.013)	(0.026)	(0.032)	(0.022)
Were days work/school missed due to illness? (1=yes)	0.797	0.004	-0.005	-0.003	0.015
	(0.402)	(0.013)	(0.013)	(0.015)	(0.020)
Waiting time before being attended (min)	4.594	0.011	-0.009	-0.011	0.044
	(1.324)	(0.016)	(0.058)	(0.056)	(0.044)
Has visited traditional health practitioner? (1=yes)	0.093	-0.021+	-0.012	-0.004	0.013
	(0.290)	(0.012)	(0.009)	(0.012)	(0.011)
Health Index	0.000	0.007	0.017	-0.034+	0.010
	(0.379)	(0.015)	(0.014)	(0.018)	(0.016)
Number of observations	12545	3673	3673	3673	2095

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the subcounty 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.

68 percent of households reported this to be the case. 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.

We also look at how the school is managed, and how stakeholders are involved. For instance, we look at whether the school has a School Management Committee (SMC) (baraza.E1.10 for primary schools, baraza.E2.10 for secondary schools). During baseline, 77 percent of households state the primary or secondary school attended by any of their children has a SMC. However, we find that, at least during baseline, only 35 percent of households is informed about SMC meetings (baraza.E1.13 and baraza.E2.13). Finally, we ask households if an inspector had visited the school in the year before the survey (baraza.E1.18 and baraza.E2.18). We find that in the baseline, households report that 50. Table 7 reports baseline means, and provides comparisons corresponding to the four main research hypotheses.

8 Additional analysis

8.1 Heterogeneity in the timing of the intervention

The slow roll-out of the intervention over an extended period also introduces variation in the time that passed between treatment administration and endline 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 endline data collection. For the most recent barazas, there will only be a few months between treatment administration and endline 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 reduce over time as promises are forgotten and plans abandoned. We thus take an agnostic stance and estimate different models that allow for both increasing and decreasing effects as a function of time elapsed between the treatment and outcome measurement. In particular, we will interact the treatment indicator with the time elapsed between treatment and endline data collection, and with its square.

Table 8 provides a summary of such a treatment heterogeneity analysis by looking only at sector level and the overall index that was also used in the overall summary of the impact evaluation (Figure 5). For each hypothesis (subcounty baraza impact, information component, deliberation component, and administrative placement) we now report estimates for three coefficients. The first, reported as (1), is the overall effect of the treatment. In a second column and indicated as (2) we report the coefficient of the treatment indicator interacted with the time that elapsed between treatment assignment in that district or subcounty (expressed in years). Finally, in column (3) we report the coefficient

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

	mean	sc baraza	information	deliberation	level
Number of children in UPS or USE	2.498	-0.004	-0.030	-0.066	0.010
	(2.080)	(0.013)	(0.068)	(0.077)	(0.078)
Distance to public school (km)	1.421	-0.001	0.008	0.038	-0.013
	(0.756)	(0.020)	(0.029)	(0.054)	(0.049)
Has complete boundary fence (1=yes)	0.309	-0.022	0.001	-0.004	0.014
	(0.462)	(0.017)	(0.017)	(0.022)	(0.020)
Has water facility (1=yes)	0.681	-0.006	0.006	0.000	-0.006
	(0.466)	(0.017)	(0.022)	(0.025)	(0.032)
Has SMC (1=yes)	0.766	-0.009	0.008	-0.008	-0.021
	(0.423)	(0.019)	(0.017)	(0.023)	(0.013)
Is informed about SMC (1=yes)	0.352	0.015	-0.003	-0.056^{+}	-0.008
	(0.478)	(0.020)	(0.023)	(0.032)	(0.019)
Inspectors visited schools (1=yes)	0.503	0.000	-0.013	-0.014	-0.037
	(0.500)	(0.021)	(0.029)	(0.027)	(0.029)
Education Index	0.000	-0.022	0.000	-0.014	-0.010
	(0.440)	(0.018)	(0.017)	(0.026)	(0.030)
Number of observations	12545	2971	2971	2971	1781

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the subcounty 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.

Table 8: Heterogeneity in timing

	sc baraza			*information*			*deliberation*			*level*		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Agriculture index	0.023	-0.030	0.010	0.003	-0.017	0.001	0.013	-0.012	0.005	0.094*	-0.082*	0.015+
	(0.033)	(0.043)	(0.013)	(0.021)	(0.031)	(0.008)	(0.021)	(0.031)	(0.011)	(0.044)	(0.041)	(0.008)
Infrastructure index	0.031	-0.027	0.007	0.012	-0.015	0.004	0.012	0.012	-0.001	-0.008	0.034	-0.009
	(0.035)	(0.054)	(0.018)	(0.026)	(0.036)	(0.010)	(0.026)	(0.036)	(0.013)	(0.046)	(0.052)	(0.010)
Health index	0.128**	-0.168**	0.042**	-0.013	-0.095**	0.018+	0.015	-0.053	0.013	-0.054	0.072	-0.014
	(0.033)	(0.043)	(0.014)	(0.023)	(0.036)	(0.010)	(0.026)	(0.044)	(0.014)	(0.066)	(0.063)	(0.013)
Education index	0.084+	-0.077	0.021	0.021	-0.041	0.006	0.039	-0.068	0.014	-0.237**	0.214*	-0.039*
	(0.048)	(0.079)	(0.025)	(0.039)	(0.051)	(0.013)	(0.034)	(0.059)	(0.019)	(0.084)	(0.094)	(0.019)
Overall index	0.227+	-0.256	0.070	0.031	-0.154	0.031	0.069	-0.084	0.020	-0.301+	0.262+	-0.048+
	(0.131)	(0.177)	(0.052)	(0.043)	(0.115)	(0.029)	(0.049)	(0.090)	(0.028)	(0.158)	(0.149)	(0.028)

Note: First column reports differences (and standard errors below) of the information component of the baraza intervention; Column 3 reports differences (and standard errors below) of the deliberation component of the baraza intervention; standard errors are clustered at the subcounty level; **, * and + denotes significance at the 1, 5 and 10 percent levels.

of the treatment indicator interacted with the square of the time that elapsed between treatment assignment in that district or subcounty to capture potential non-linearities.

8.2 Endline imbalance and matched difference-in-difference

While the 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 change over time. Therefore, comparing characteristics between control sub-counties and sub-counties that were allocated to a treatment cell but ended up not being treated will be an important strategy to test for selection bias that may have been introduced by the incomplete roll-out of the treatment. In particular, instead of simply collecting information from the (planned) control sub-counties, we will also collect information from sub-counties that were supposed to receive a treatment but did not get one. If the incomplete roll-out introduced selection bias, this information may also be useful to assess the direction and magnitude of the bias.

Table 9 thus compares endline outcomes between households that were planned to receive a particular treatment 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; full results with details for each outcome in the four sectors can be found in Appendix 11.2.

As mentioned in Section 5.3, if we find evidence of imbalance between planned but untreated subcounties and planned control sub-counties using endline information, we will try to recover unbiased impact estimates using a matched difference-in-difference estimator. Similar to Figure 5, Figure 6 summarizes the results for the first three research questions (impact of sub-county baraza, relative importance of information component, and relative importance of deliberation component) using such an estimator. In Appendix 11.3, full results are reported.

9 Conculsion

To improve governance and 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. In 2015, we designed a study aimed at evaluating the effectiveness of community advocacy forums, also known as barazas, in Uganda. 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) how does a baraza organized at the district level compare to one organized at the sub-county level. Baseline data

Table 9: Difference between planned but not treated subcounties and planned controls at endline

	sc baraza	information	deliberation
Agriculture index	0.010	0.007	-0.004
	(0.023)	(0.019)	(0.022)
Infrastructure index	0.005	0.022	0.004
	(0.033)	(0.021)	(0.021)
Health index	-0.017	0.010	-0.009
	(0.015)	(0.013)	(0.015)
Education index	0.022	0.015	0.016
	(0.023)	(0.018)	(0.020)
Public service delivery index	0.009	0.032	0.003
	(0.032)	(0.020)	(0.023)
Number of observations	1995	2891	3440

Note: First column reports differences (and standard errors below) of the information component of the baraza intervention; Column 3 reports differences (and standard errors below) of the deliberation component of the baraza intervention; standard errors are clustered at the subcounty level; **, * and + denotes significance at the 1, 5 and 10 percent levels.

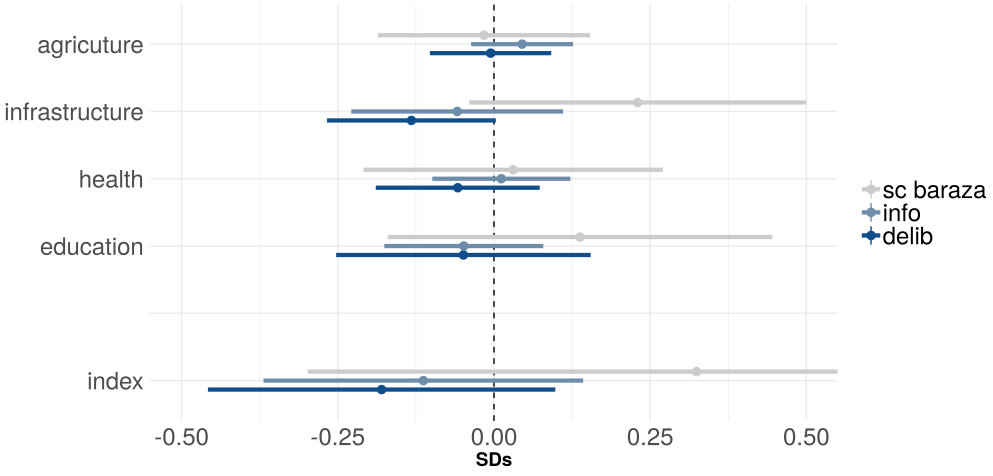


Figure 6: Summary of baraza impact (matched difference in difference)

on more than 12,500 households spread over almost 250 sub-counties in about 40 districts throughout Uganda was collected and OPM stared implementing barazas following our protocol.

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 endline information after partial roll-out. Various strategies were followed to test, and reduce the consequences of, potential selection bias introduced by this partial roll out. Analysis of the baseline data in light of the partial roll-out, documented in this report, do not suggest that the roll-out introduced bias.

This pre-registered report serves as a dummy report for the yet to be collected endline data. Currently, most of the tables are populated using simulated data. However, the purpose of developing and publishing this detailed report in this way is to tie our hands: it will reduce our researcher degrees of freedom in the selection of what specifications to run and what variables to select to assess impact of the baraza (Simmons, Nelson, and Simonsohn, 2011). In an effort to be as transparent as possible, this entire project with all code is publicly available from github. Endline data collection is scheduled to start on January 17th, 2020. We hope to be able to compile a version of this report with endline data as soon as endline data collection is finished (which is scheduled to be the beginning of March).

10 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 (https://www.cgiar.org/funders/) under grant number PIM 203002.002.221. 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. 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.

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.

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.

Gilens, M. 2001. "Political Ignorance and Collective Policy Preferences." *The American Political Science Review* 95 (2): 379–396.

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

Golooba-Mutebi, F. 2005. "When popular participation won't improve service provision: primary health care in Uganda." *Development Policy Review* 23 (2): 165–182.

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.

Iacus, S. M., G. King, and G. Porro. 2012. "Causal inference without balance checking: Coarsened exact matching." *Political analysis* 20 (1): 1–24.

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.

Simmons, J. P., L. D. Nelson, and U. Simonsohn. 2011. "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant." *Psychological science* 22 (11): 1359–1366.

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.

Figure A.1: MDEs Agricultural Sector Outcomes

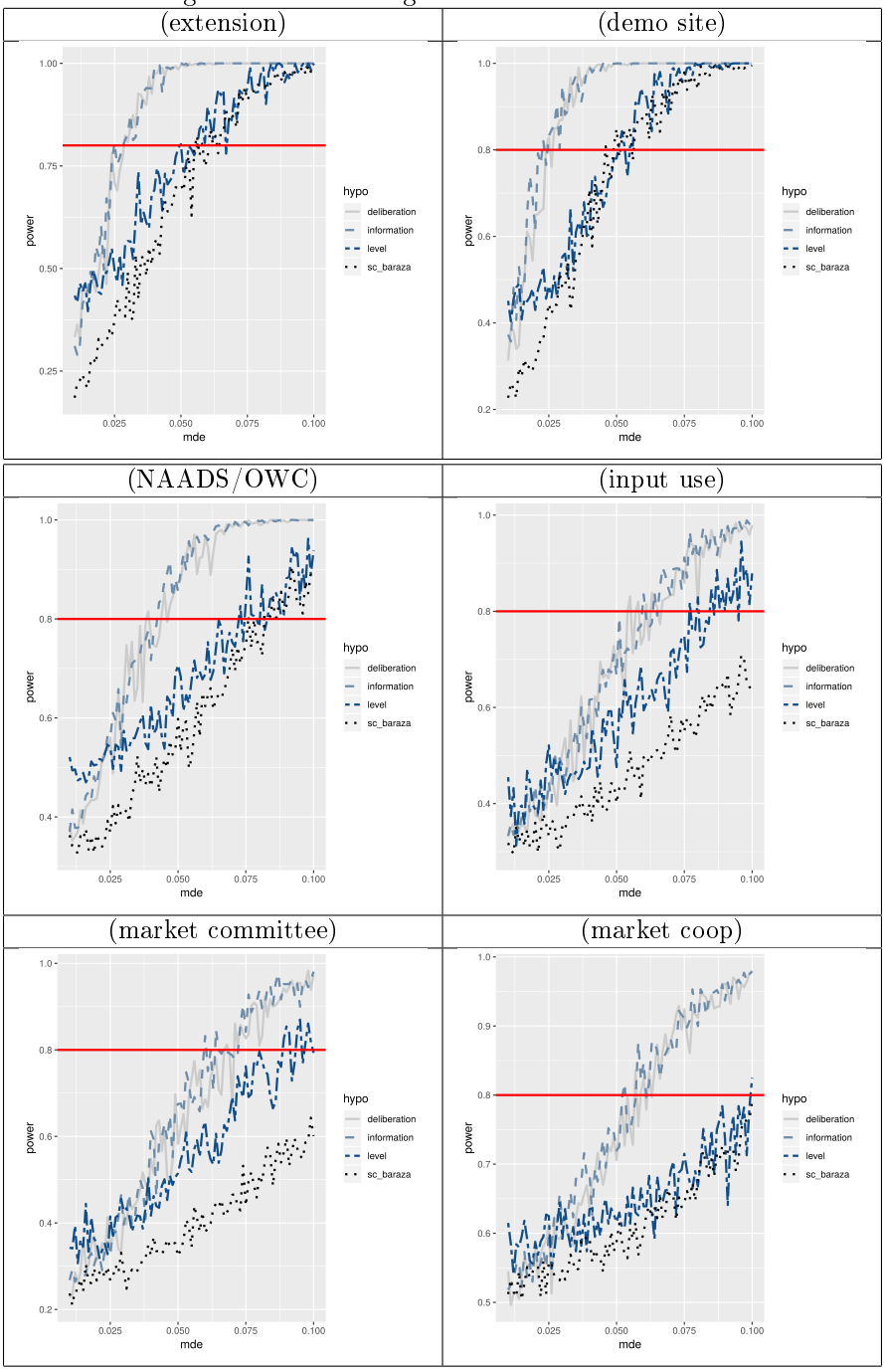
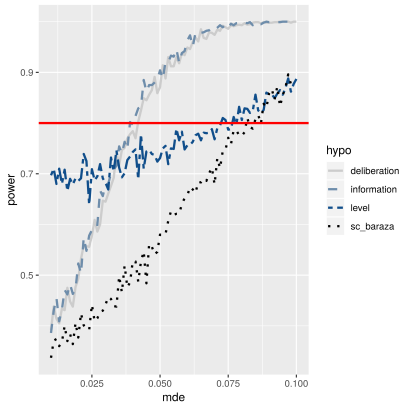


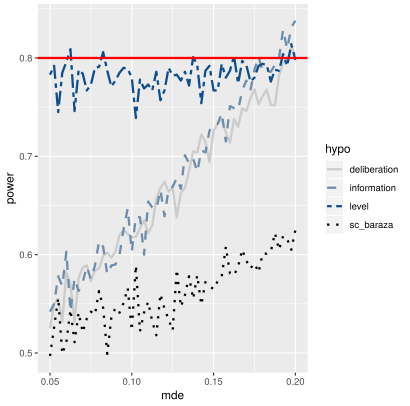
Figure A.2: MDEs Infrastructure Outcomes

(unprotected source)

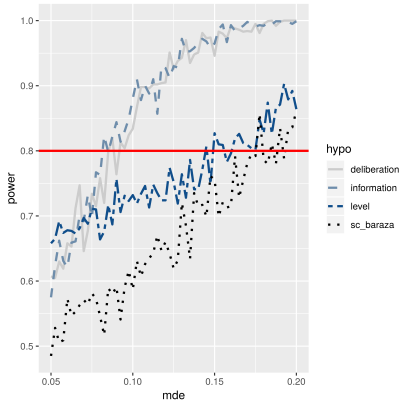
(distance source)



(waiting time source)



(Water User Committee)



(distance road)

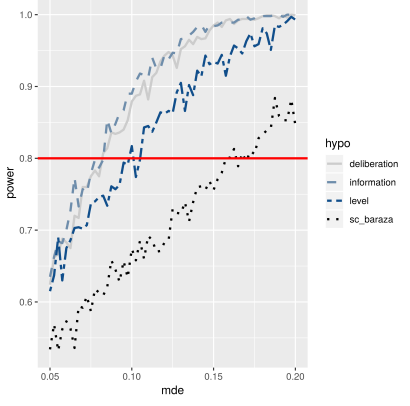


Figure A.3: MDEs health sector outcomes

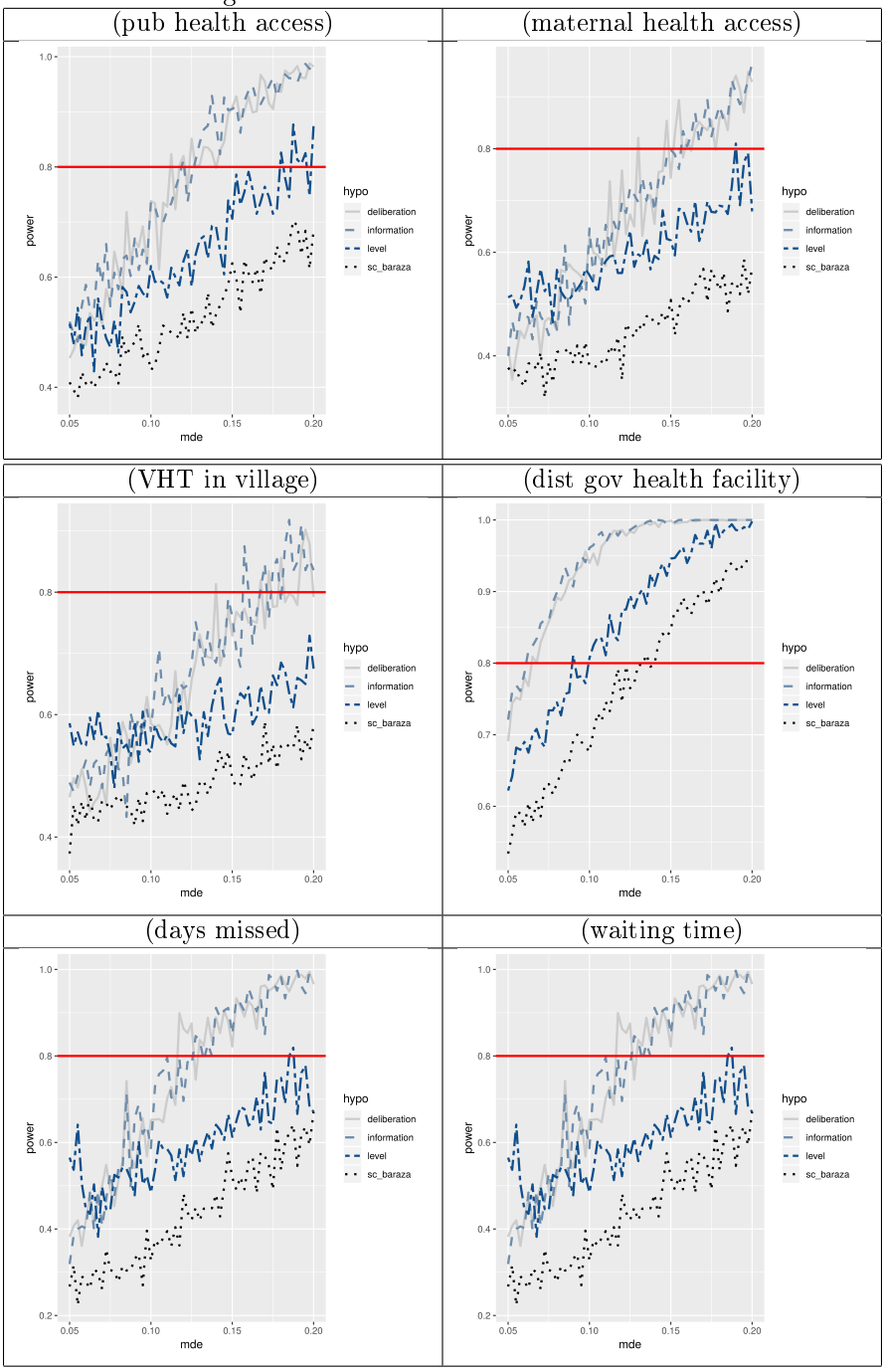
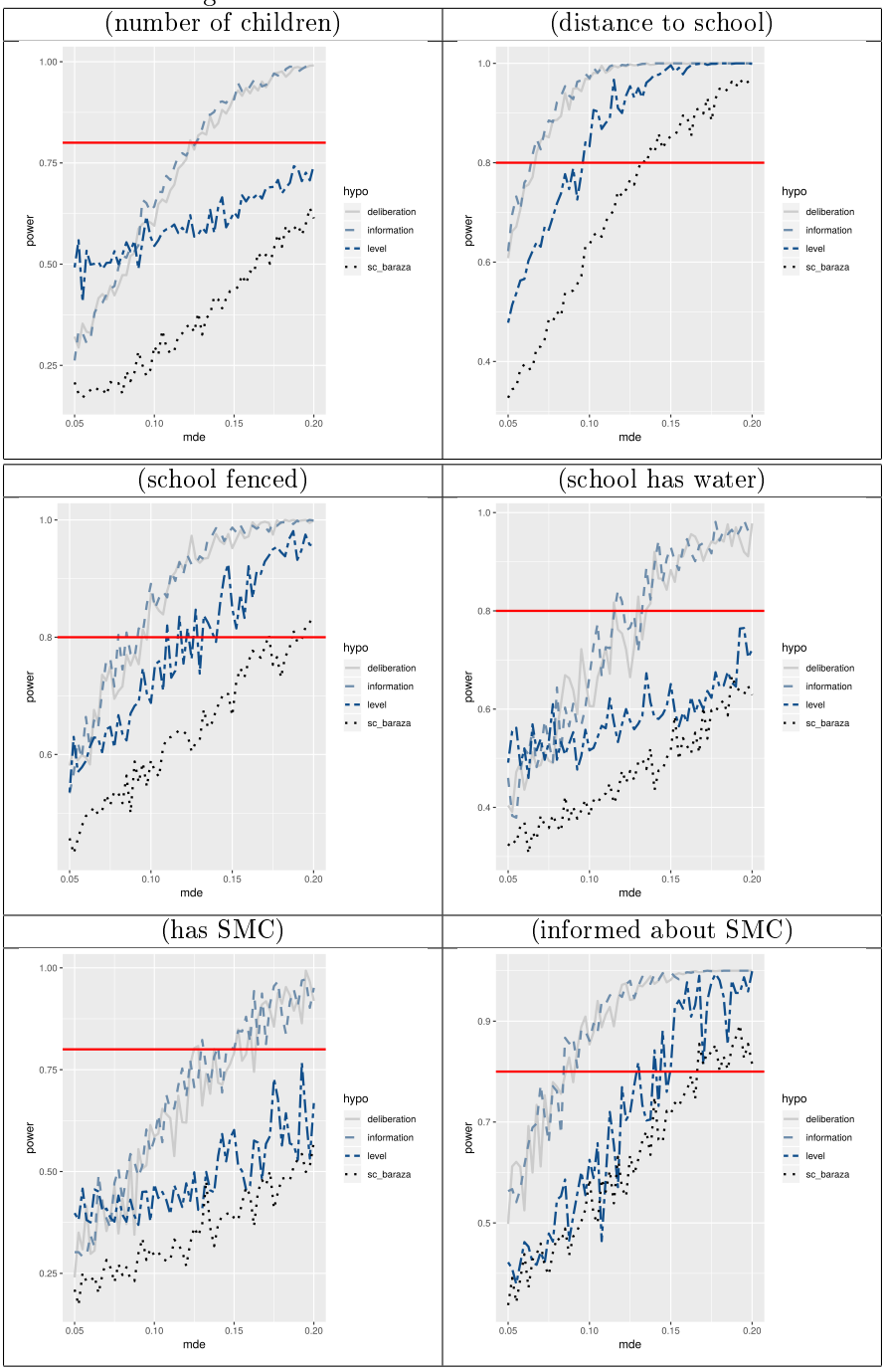


Figure A.4: MDEs education sector outcomes



11 Appendices

11.1 Appendix one – updated power calculations

11.2 Appendix Two – balance between planned but untreated sub-counties and planned control sub-counties

11.3 Appendix Three – Results for matched difference-in-difference models

Table A.1: Difference between planned but not treated subcounties and planned control subcounties – Agriculture

	sc baraza	information	deliberation
Was visited by extension officer at home (yes/no)	-0.005	0.003	0.011
	(0.014)	(0.012)	(0.016)
Visited training or demonstration site (yes/no)	0.008	0.006	-0.021
	(0.013)	(0.011)	(0.014)
NAADS or OWC in village (yes/no)	0.004	-0.010	-0.001
	(0.016)	(0.014)	(0.018)
Uses modern inputs (improved seed or fertilizer) (yes/no)	-0.009	0.012	0.022
	(0.022)	(0.017)	(0.024)
Support in marketing from village procurement committe (yes/no)	0.010	0.009	0.023
	(0.024)	(0.020)	(0.023)
Support in marketing from cooperative (yes/no)	0.015	-0.003	-0.031
	(0.030)	(0.022)	(0.029)
Agriculture index	0.010	0.007	-0.004
	(0.023)	(0.019)	(0.022)
Number of observations	1995	2891	3440

Note: First column reports differences (and standard errors below) of the information component of the baraza intervention; Column 3 reports differences (and standard errors below) of the deliberation component of the baraza intervention; standard errors are clustered at the subcounty level; **, * and + denotes significance at the 1, 5 and 10 percent levels.

Table A.2: Difference between planned but not treated subcounties and planned control subcounties – Infrastructure

	sc baraza	information	deliberation
Household uses unprotected water source during dry season (yes/no)	-0.019	-0.038*	0.011
	(0.023)	(0.017)	(0.020)
Distance to water source	-0.011	0.006	-0.021
	(0.033)	(0.021)	(0.028)
Average waiting time at source (min)	0.125	0.020	-0.092
	(0.140)	(0.100)	(0.120)
Is there a Water User Committee in this village? (yes/no)	-0.004	0.003	-0.007
	(0.032)	(0.022)	(0.025)
Distance to nearest all weather road (km)	0.025	0.033	0.058
	(0.051)	(0.034)	(0.039)
Infrastructure index	0.005	0.022	0.004
	(0.033)	(0.021)	(0.023)
Number of observations	1995	2891	3440

Note: First column reports differences (and standard errors below) of the information component of the baraza intervention; Column 3 reports differences (and standard errors below) of the deliberation component of the baraza intervention; standard errors are clustered at the subcounty level; **, * and + denotes significance at the 1, 5 and 10 percent levels.

Table A.3: Difference between planned but not treated subcounties and planned control subcounties – Health

	sc baraza	information	deliberation
Seek treatment for fever in public health facility (1=yes)	-0.020	0.007	0.001
	(0.020)	(0.017)	(0.023)
Go to public health facility to give birth (1=yes)	0.020	0.028*	0.012
	(0.012)	(0.011)	(0.017)
Is there a VHT in village? (1=yes)	0.006	-0.008	-0.017
	(0.019)	(0.016)	(0.017)
Distance to nearest govt health facility (km)	0.044	-0.019	0.033
	(0.027)	(0.025)	(0.025)
Were days work/school missed due to illness? (1=yes)	0.011	-0.012	-0.022
	(0.018)	(0.013)	(0.015)
Waiting time before being attended (min)	0.129*	0.090*	0.066
	(0.054)	(0.042)	(0.048)
Has visited traditional health practitioner? (1=yes)	-0.012	-0.004	0.007
	(0.017)	(0.012)	(0.011)
Health index	-0.017	0.010	-0.009
	(0.015)	(0.013)	(0.015)
Number of observations	1995	2891	3440

Note: First column reports differences (and standard errors below) of the information component of the baraza intervention; Column 3 reports differences (and standard errors below) of the deliberation component of the baraza intervention; standard errors are clustered at the subcounty level; **, * and + denotes significance at the 1, 5 and 10 percent levels.

Table A.4: Difference between planned but not treated subcounties and planned control subcounties – Education

	information	deliberation	
Number of children in UPS or USE	0.074	0.050	-0.004
	(0.095)	(0.080)	(0.072)
Distance to public school (km)	0.015	0.028	0.012
	(0.034)	(0.032)	(0.042)
Has complete boundary fence (1=yes)	0.034	0.006	-0.011
	(0.023)	(0.018)	(0.027)
Has water facility (1=yes)	0.016	0.020	0.038
	(0.016)	(0.017)	(0.024)
Has School Management Committee (1=yes)	-0.006	-0.005	0.015
	(0.018)	(0.014)	(0.027)
Is informed about School Management Committee (1=yes)	0.023	0.029	0.007
	(0.030)	(0.020)	(0.024)
Inspectors visited schools (1=yes)	-0.012	0.029	-0.011
	(0.028)	(0.025)	(0.023)
Education index	0.022	0.015	0.016
	(0.023)	(0.018)	(0.020)
Number of observations	1995	2891	3440

Note: First column reports differences (and standard errors below) of the information component of the baraza intervention; Column 3 reports differences (and standard errors below) of the deliberation component of the baraza intervention; standard errors are clustered at the subcounty level; **, * and + denotes significance at the 1, 5 and 10 percent levels.

Table A.5: Treatment-control differences (matched difference-in-difference) - Agriculture

	mean	sc baraza	information	deliberation	level
Was visited by extension officer at home (yes/no)	0.112	-0.007	0.023	0.016	-0.017
	(0.316)	(0.050)	(0.026)	(0.023)	(0.030)
Visited training or demonstration site (yes/no)	0.080	-0.010	0.006	-0.011	-0.014
	(0.271)	(0.050)	(0.027)	(0.026)	(0.025)
NAADS or OWC in village (yes/no)	0.179	0.039	0.008	-0.012	-0.056
	(0.384)	(0.068)	(0.035)	(0.035)	(0.037)
Uses modern inputs (improved seed or fertilizer) (yes/no)	0.339	0.033	0.000	-0.022	0.059
	(0.473)	(0.094)	(0.062)	(0.052)	(0.052)
Support in marketing from village procurement committe (yes/no)	0.439	-0.039	0.048	0.029	-0.015
	(0.496)	(0.080)	(0.051)	(0.047)	(0.049)
Support in marketing from cooperative (yes/no)	0.269	-0.015	0.020	-0.031	-0.019
	(0.444)	(0.114)	(0.064)	(0.065)	(0.060)
Agriculture Index	0.000	-0.016	0.045	-0.005	-0.046
	(0.409)	(0.085)	(0.040)	(0.048)	(0.039)
Number of observations	12545	2846	4016	4060	3844

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the subcounty 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.

Table A.6: Treatment-control differences (matched difference-in-difference) - Infrastructure

	mean	sc baraza	information	deliberation	level
Household uses unprotected water source during dry season (yes/no)	0.696	-0.070	0.048	-0.094	0.014
	(0.460)	(0.136)	(0.082)	(0.102)	(0.101)
Distance to water source	0.708	-0.019	-0.051	-0.014	0.057
	(0.528)	(0.130)	(0.078)	(0.067)	(0.123)
Average waiting time at source (min)	2.712	-1.499*	0.140	1.557**	0.078
	(2.169)	(0.702)	(0.374)	(0.346)	(0.599)
Is there a Water User Committee in this village? (yes/no)	0.525	0.024	-0.037	0.044	-0.118
	(0.499)	(0.109)	(0.076)	(0.059)	(0.090)
Distance to nearest all weather road (km)	0.896	0.079	0.102	0.084	-0.085
	(0.912)	(0.352)	(0.210)	(0.210)	(0.118)
Infrastructure Index	0.000	0.230+	-0.059	-0.132*	-0.096
	(0.449)	(0.134)	(0.084)	(0.066)	(0.067)
Number of observations	12545	2422	3540	3480	3284

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the subcounty 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.

Table A.7: Treatment-control differences (matched difference-in-difference) - Health

	mean	sc baraza	information	deliberation	level
Seek treatment for fever in public health facility (1=yes)	0.736	-0.058	-0.028	0.048	-0.009
	(0.441)	(0.134)	(0.061)	(0.071)	(0.057)
Go to public health facility to give birth (1=yes)	0.840	-0.091	-0.001	0.062	0.024
	(0.366)	(0.080)	(0.039)	(0.046)	(0.065)
Is there a VHT in village? (1=yes)	0.752	-0.037	0.018	-0.072	-0.123*
	(0.432)	(0.088)	(0.060)	(0.045)	(0.061)
Distance to nearest govt health facility (km)	1.998	-0.039	-0.050	0.009	-0.024
	(0.745)	(0.225)	(0.126)	(0.150)	(0.123)
Were days work/school missed due to illness? (1=yes)	0.797	0.008	-0.008	-0.006	0.005
	(0.402)	(0.085)	(0.040)	(0.047)	(0.039)
Waiting time before being attended (min)	4.594	0.106	0.071	0.196	0.178
	(1.324)	(0.259)	(0.144)	(0.142)	(0.142)
Has visited traditional health practitioner? (1=yes)	0.093	0.069	-0.026	-0.010	-0.002
	(0.290)	(0.053)	(0.028)	(0.025)	(0.032)
Health Index	0.000	0.031	0.012	-0.058	-0.064
	(0.379)	(0.119)	(0.055)	(0.064)	(0.073)
Number of observations	12545	1744	2422	2462	2416

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the subcounty 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.

Table A.8: Treatment-control differences (matched difference-in-difference) - Education

	mean	sc baraza	information	deliberation	level
Number of children in UPS or USE	2.498	-0.221	0.114	0.234	-0.136
	(2.080)	(0.324)	(0.194)	(0.168)	(0.214)
Distance to public school (km)	1.421	0.206	0.066	0.077	-0.166^{+}
	(0.756)	(0.269)	(0.097)	(0.145)	(0.099)
Has complete boundary fence (1=yes)	0.309	0.218	0.045	-0.040	0.085
	(0.462)	(0.176)	(0.075)	(0.105)	(0.103)
Has water facility (1=yes)	0.681	0.214^{+}	-0.096	-0.071	-0.056
	(0.466)	(0.129)	(0.061)	(0.076)	(0.072)
Has SMC (1=yes)	0.766	0.209^{+}	-0.047	-0.056	-0.102*
	(0.423)	(0.118)	(0.065)	(0.061)	(0.045)
Is informed about SMC (1=yes)	0.352	0.065	0.048	0.070	-0.077
	(0.478)	(0.140)	(0.094)	(0.075)	(0.058)
Inspectors visited schools (1=yes)	0.503	0.291	0.002	-0.048	-0.062
	(0.500)	(0.186)	(0.087)	(0.076)	(0.077)
Education Index	0.000	0.138	-0.048	-0.049	0.018
	(0.440)	(0.151)	(0.063)	(0.099)	(0.071)
Number of observations	12545	1104	1612	1732	1754

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the subcounty 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.