Community advocacy forums and public service delivery in Uganda: impact and the role of information, deliberation, and administrative placement

Bjorn Van Campenhout
International Food Policy Research Institute, Belgium

Nassul Kabunga
Uganda Bureau of Statistics, Uganda

Tewodaj Mogues
International Monetary Fund, USA

Caroline Miehe
KU Leuven, Belgium

Grantee Final Report

Accepted by 3ie: March 2021

Note to readers

This impact evaluation has been submitted in fulfilment of requirements under grant PW2.15 issued under Policy Window 2. This version is being published online as it was received. A copy-edited and formatted version will be available in the 3ie Impact Evaluation Report Series in the near future.

All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors and omissions are the sole responsibility of the authors. All affiliations of the authors listed in the title page are those that were in effect at the time the report was accepted. Any comments or queries should be directed to the corresponding author, Bjorn Van Campenhout, at: b.vancampenhout@cgiar.org

The 3ie technical quality assurance team comprises Sayak Khatua, Francis Rathinam, an anonymous external impact evaluation design expert reviewer and an anonymous external sector expert reviewer, with overall technical supervision by Sebastian Martinez.

Funding for this impact evaluation was provided by Bill & Melinda Gates Foundation. A complete listing of all of 3ie's donors is available on the 3ie website.

Suggested citation: Campenhout, BV, Kabunga, N, Mogues, T and Miehe, C, 2021. *Community advocacy forums and public service delivery in Uganda: impact and the role of information, deliberation, and administrative placement,* 3ie Grantee Final Report. New Delhi: International Initiative for Impact Evaluation (3ie).

Acknowledgements

This study 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 Consultative Group on International Agricultural Research (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.

Executive summary

Introduction

Ineffective monitoring and weak accountability mechanisms impaired Uganda's development since its independence. In 2009, the Government of Uganda initiated community advocacy forums, also known as barazas, to involve the public in holding the government accountable for its performance in relation to the resources spent and to finally improve public service delivery. The baraza programme was initiated by the president of Uganda and implemented by the Office of the Prime Minister (OPM).

We proposed a cluster randomized control trial to evaluate these barazas. This study's overall purpose was to establish, in a rigorous way, if the program had an impact on public service delivery. A second objective of the study was to inform policy makers about the relative effectiveness of barazas organized at lower administrative levels (the sub-county) to that of barazas organized at a more aggregate level (the district). Third, the study also set out to explore pathways through which community advocacy forums may affect outcomes. Using a two-by-two factorial design, it differentiates between (1) the impact of providing citizens with information, and (2) the impact of letting citizens engage with public servants and politicians.

Intervention

To achieve these objectives, we designed four interventions: sub-county level barazas were implemented at the sub-county level and included information and deliberation. To study the relative importance of both components, we used this baraza and either removed the information component or the deliberation component. District level baraza were similar to sub-county level barazas (including information and deliberation), but were organized at district level. We trained local government officials to ensure adherence to our intervention protocols, and the interventions were rolled out by the OPM, our main implementing partner.

Evaluation questions, methods, design, sampling and data collection

These interventions correspond to the following evaluation questions: What is the impact of sub-county level barazas on public service delivery? What is the relative importance of the information component and of the deliberation component? What is the impact of district level barazas on public service delivery?

A baseline survey with more than 12,500 households and 400 government officials was conducted in 2015. While the study was initially assumed to take 2 years, OPM faced various implementation challenges. 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 and conducting the end-line survey after partial roll-out. We decided to collect end-line data and employ estimation and data collection strategies to control for potential selection bias. In 2020, end-line data was collected on 6,700 households and 260 government officials.

In a first part of this study, we strictly follow a pre-analysis plan that summarizes a range of outcomes corresponding to four main sectors (agriculture, infrastructure, health and education) as indices and one overall index of public service delivery. In a second part of the study, we proceed in a more exploratory way. We provide a detailed analysis of individual outcomes. We look at each of the sectors in more detail and for changes in behaviour that is explicitly targeted

by barazas. We also provide results on changes in the perception of citizens on a range of issues. In a third part, we explore heterogeneity in the treatment effect.

Findings

Judging by the pre-analysis plan's summary indices in the first part, we find little evidence that the baraza intervention had an impact on public service delivery. The only exception is agriculture, where sub-county level barazas have a positive impact, and where this impact is superior to the (lack of) impact associated with district level barazas. The second and third part add more nuance to this conclusion. For instance, we find that in the agricultural sector, sub-county level barazas significantly increase access to agricultural extension, a common practice to transmit agricultural information and technologies to farmers. However, this seems to be driven by households that live close to the sub-county headquarter. Looking at infrastructure outcomes, we find that sub-county level barazas reduce waiting time at the water source. Interestingly, this effect seems to be strongest in more remote areas. For health, we only find effects if we restrict the sample to sub-counties where officials recall that a baraza happened; then we find that the information and deliberation components affect the use of government health facilities. For education, we see an increase in enrolment rates, but only if enough time has passed between the intervention and the end-line data collection.

Cost analysis

The complex picture that emerges from this analysis also means that conclusions in terms of cost-effectiveness are not unambiguous. For instance, with regard to public service delivery related to water infrastructure, district level barazas are far more cost effective than sub-county level barazas, as many more households can be reached. However, with respect to agriculture, sub-county level barazas are most cost-effective. The same holds for comparisons between the cost-effectiveness of the deliberation and information component. In general we find that since baraza interventions impact large numbers of households and cost relatively little, the rate of return is substantial, even if treatment effects are small in size.

Discussion

These mixed results are puzzling, especially because prior qualitative research suggested an effect of the intervention (Van Campenhout et al. 2018). We suspect that the lack of quantifiable impact can be explained by the nature of the intervention. Barazas address various issues in heterogeneous settings: different sub-counties face different challenges, so that different issues are discussed and prioritized during the barazas. The actual baraza treatments may thus be far from standardized and their impact may be highly localized and context specific. As a result, a focus on average treatment effects may fail to identify a significant impact. The effect is attenuated because it is averaged over many sub-counties that in reality received "different" types of barazas. While it is good that barazas tend to focus on and potentially affect areas that are most problematic, this complicates the estimation and might be the reason why we find only limited effects. Hence, barazas might work while we are unable to detect this. Concerns related to non-standardized treatments are confirmed when looking at heterogeneous treatment effects and a case study of access to water in Bagezza sub-county. That is why we recommend baraza meetings even though they do not have a measurable effect on our pre-registered indices. We suggest a mix of district level and sub-county level barazas and recommend the implementation of full barazas that are held several times, for instance every two years.

Contents

List of figures and tables

Abbreviations and acronyms

AIDS	Acquired immune deficiency syndrome
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
ASSP	Agriculture sector strategic plan
CAO	Chief administrative officer
CGIAR	Consultative Group on International Agricultural Research
CRSE	Cluster robust standard errors
DAP	Diammonium phosphate
e.g.	Exempli gratia, for example
GoU	Government of Uganda
GPS	Global positioning system
HC2	Health centre 2
HC3	Health centre 3
HH	Household
HIV	Human immunodeficiency viruses
HQ	Headquarter
i.e.	Id est, in other words
IFPRI	International Food Policy Research Institute
IHS	Inverse hyperbolic sine
km	Kilometre
LC1	Village local council chairperson
LC3	Sub-county local council chairperson
LC5	District local council chairperson
MAKSS REC	Makerere university school of social sciences research ethics committee
MDE	Minimal detectable effects
MOP	Muriate of potash
min	Minutes
NAADS	National agricultural advisory services
NGO	Non-governmental organization

NPK	Nitrogen, phosphorus and potassium
ODK	Open data kit
OPM	Office of the prime minister
OWC	Operation wealth creation
PIM	Policies, institutions, and markets
PTA	Parent teacher association
RCT	Randomized controlled trial
RDC	Resident district commissioner
SC	Sub-county
SDG	Sustainable development goals
SMC	School management committee
SSP	Single superphosphate
TSP	Triple superphosphate
UGX	Uganda shillings
UNCST	Uganda national council for science and technology
UPE	Universal primary education
US	United States
USE	Universal secondary education
VHT	Village health teams
3ie	International Institute for Impact Evaluation

1. Introduction

Since Uganda's independence in 1962, the country's development efforts have been thwarted by political turmoil and economic mismanagement. In the mid-1980s, after attainment of relative stability, the Government of Uganda (GoU) supported by development partners, initiated reforms to address development challenges of the time. Notable among these initiatives was the liberalization of the economy and the introduction of a decentralized system of governance (Francis and James 2003, Benin et al. 2007). Decentralization was particularly viewed as a suitable mechanism for addressing welfare and political challenges by improving efficiency of public service delivery, formulating more appropriate services, bringing representative governance closer to citizens (Steiner 2007, Francis and James 2003). A major ingredient of decentralization is to enhance empowerment and build a sense of ownership of the local citizens to actively participate in planning, implementation and evaluation of development interventions in their locations, to improve accountability and responsiveness of local leaders and service providers (Burki et al. 1999).

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

As barazas continued to be rolled out beyond the pilot communities, a rigorous evaluation of their effectiveness was still outstanding. This study's overall purpose is to establish, in a rigorous way, if the program had an impact on public service delivery. A second objective of the study is to inform policy makers about the relative effectiveness of barazas organized at lower administrative levels to that of barazas organized at a more aggregate level. Third, the study also set out to explore pathways through which community advocacy forums may affect outcomes, as we differentiate between the impact of providing citizens with information, and the impact of letting citizens engage with public servants and politicians. At the time of the proposal, the Government of Uganda shared the same aspiration so as to inform policy on program effects of service delivery to local communities (OPM 2013) since there had not been any formal study conducted to test the actual achievements of the baraza initiative so far against the

[1] The initial pilot barazas were undertaken in eight lower level local governments (generically referred to as sub-counties) 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.

set objectives. From a policy perspective, it was also important to assess if the switch from sub-county level barazas to district level barazas was cost-effective. Up to date, the OPM has been eager to learn about the results from the different components of the impact evaluations and (preliminary) results have been presented at various high-level meetings.

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 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 modelled on the one of Björkman and Svensson (2009). 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 run.

The 3ie Systematic Review 43 (2019) also discusses whether citizens engagement in the planning, management and oversight of public services affects the quality of and access to those services and citizens' quality of life. In some programmes, citizens participate in setting the priorities for and the planning of local services (Touchton and Wampler 2014, Goncalves 2013, Diaz-Cayeros et al. 2014, Beuermann and Amelina 2014, Ananthpur et al. 2014, Giné et al. 2018, Humphreys et al. 2014, Beath et al. 2013). Other programmes evaluate transparency mechanisms, aimed to disclose and disseminate information, such as public official or service provider performance information interventions (Humphreys and Weinstein 2012, Grossman and Michelitch 2018, Timmons and Garfias 2015, Capuno and Garcia 2010). Moreover, evaluations of accountability mechanisms are included, which comprised citizen feedback or monitoring mechanism interventions to hold public service providers and institutions responsible for executing their powers and mandates according to appropriate standards (Berman et al. 2017, Alhassan et al. 2016, Grossman et al. 2017, Björkman et al. 2009, Björkman et al. 2017, Gullo et al. 2017, Bradley and Igras 2005, Molina 2014). This review shows that interventions promoting citizen engagement by improving direct engagement between service users and service providers are often effective in stimulating citizen engagement and in improving public service delivery but complementary interventions that address bottlenecks around service provider supply chains and service use are needed to improve wellbeing. On the other hand, interventions promoting citizen engagement by increasing citizen pressures on politicians to hold service providers accountable and thus improve governance often do not influence service delivery.

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 consisting of only a few villages.

Second, we explore pathways through which community advocacy forums may affect public service delivery. A two-by-two factorial design enables us to differentiate between (1) the impact of providing citizens with information related to budgeting and planning, and (2) the impact of letting citizens engage with public servants and politicians in a facilitated questions-and-answers session. On the one hand, informational interventions can increase political accountability (Dunning et al. 2019). A citizen who is informed about the performance of politicians and civil servants, can monitor the latter and apply pressure (Raffler, Posner, Parkerson 2018). There is some evidence that providing citizens with information about public services can increase their ability to hold leaders accountable to improve public service delivery (Pandey, Goyal, Sundararaman, 2009, Gilens, 2001). On the other hand, deliberation can increase the quality of public services as well. Citizens can confront their leaders with urgent and important matters and threaten them if they are not performing. Creating a platform where stakeholders can meet and interact may also increase mutual understanding and result in a better relationship between them. The impact of deliberation has also been the subject of empirical analysis (Björkman Nyqvist, de Walque, Svensson 2017, Goeree and Yariv 2011, Fujiwara and Wantchekon 2013).

Third, 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 non-governmental organizations (NGOs). 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 or education. Furthermore, it is likely to be easier to re-allocate resources to problems identified during barazas if they are organized by the government. This is also consistent with suggestive evidence in Raffler, Posner, and Parkerson (2018), who find that the presence of sub-county officials during their community-based monitoring intervention boosted the impact of the intervention. However, effects may also work in the opposite direction. For example, an intervention to reduce absenteeism in government public health facilities in India was initially very successful but ceased to have any impact after the local bureaucracy started providing official excuses for most of the nurses' absences (Banerjee, Duflo, Glennerster 2008). Most of the other studies that are closest to our study partnered with NGOs for implementation (for example (e.g.) Björkman and Svensson 2009, Raffler, Posner, Parkerson 2018).

Fourth, 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 as hygiene in health centers or accessibility, involve cooperation between heads of different sectors (e.g. health and infrastructure to get access to water in health centers or access roads). 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 existing studies focus on a single sector; the health sector in particular seems to be a popular sector for community monitoring interventions (e.g. Arkedis et al. 2019, Björkman and Svensson 2009, Raffler, Posner, Parkerson 2018).

Finally, we evaluate a high-profile policy intervention that receives broad support within government and among citizens alike in Uganda. Evaluating policy interventions has it challenges, and this one is no exception. As a result, such research has become rare, as present day randomized controlled trial (RCT) often bypass the political resistance to randomization among governments, development workers and beneficiaries, as the nature of the partners has changed (NGOs rather than governments) and the interventions have become “relatively trivial” (de Souza Leão and Eyal 2019).

In this study, we start by providing a brief overview of the government program we evaluated and explain the theory of change behind the components of the intervention. We then present the four main research questions and provide details on the cluster randomized control trial we used to answer these questions. This section also provides information on the sampling frame and presents detailed power simulations that account for the consequences of the implementation challenges. This is followed by an explanation on how the implementation deviated from what was planned and the strategies that were used to diagnose and remedy the potential bias introduced by this deviation We then present the findings, starting with balance tables and results of a pre-registered analysis. We provide further details and also look at outcomes that were not pre-registered to explore some of the mechanisms behind the intervention. This part also includes an extensive analysis of sub-county level data that was collected from government officials. We then present heterogeneous treatment effects and reflect on the partial roll-out as a threat to study validity. We also provide a cost-benefit analysis. The penultimate section provides a discussion of the results and a final section concludes.

2. Intervention

2.1 Description

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 policy makers and civil servants 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.

A typical baraza is initiated from the center, with the OPM mobilizing district and sub-county officials. These include the Chief Administrative Officer (CAO) as the head of public service 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 sub-county level, the sub-county level equivalents of the CAO (the sub-county chief) and the LC5, the sub-county local council 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 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 sub-county (or the entire district in case of a district level baraza); 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.

In our study, we do not only want to test if barazas work. We also want to learn which of the main components – the deliberation component or the information component – are responsible for most of the effect. Finally, we also want to directly compare the effectiveness of district level barazas to that of sub-county level barazas. We thus differentiate between four types of barazas: a sub-county level baraza, an information baraza, a deliberation baraza and a district level baraza.

The sub-county level barazas are basically the barazas as they were implemented by the OPM at the sub-county level. They have both an information and deliberation component. To study the relative importance of the information component and the deliberation component respectively, we used this baraza as a starting point and, either removed the information component or the deliberation component from the generic sub-county level baraza to test the relative importance of these components.

The information component of a baraza involves templates that were developed to be filled by officials and mounted at a central location in each parish of the district two weeks before the baraza. The template was designed to inform citizens about planned and actual public expenditures for the previous fiscal year, about achievements and challenges encountered during that year, and about planned expenditures and targets for the next fiscal year. This needed to be filled for each of the four sectors (agriculture, infrastructure, health and education) by the sub-county chief.

On the day of the baraza event, the CAO provided a brief presentation on overall budget/finances for the fiscal year, main achievements and challenges in service delivery, and introduced local officials. After a brief intervention by the OPM, local officials responsible for each sector then presented more or less the same as what was required for the templates. An information focused baraza allowed for only 10 clarifying questions to be asked, to be collected and asked by the facilitator.

For the deliberation component of the barazas, posters were also mounted in each parish of the sub-county, but only to announce that a baraza will be held at a particular date and place. At the baraza itself, after a brief introduction by the RDC, citizens are guided to break into 5 groups by sector, discuss problems they face and draw up a list of priority issues that need to be addressed. Facilitators in each group are required to anonymously collect these issues and concerns. Facilitators are expected to focus the discussion on what was done well, and what were the problems during the past year. The discussions should also result in agreement on what should be done in the next fiscal year. After the break-out sessions, officials are asked to react to the specific comments and requests.

District level barazas were very similar to sub-county level barazas (that is, with both an information and deliberation component), except for the fact that district level barazas are organized at the district headquarters and all sub-county chiefs and sub-county chairpersons (LC3's) of each sub-county within the district are expected to attend in case issues arise related to their sub-county.

2.2 Theory of Change

- The impact of (sub-county level) barazas

The baraza intervention fundamentally seeks to improve public services through improving accountability of local public decision makers and service providers. The baraza intervention as conceived by the OPM is a fairly standard community-based monitoring intervention that combines the provision of information with the possibility of citizens to engage with each other and with decision makers and public servants 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). As the subcounty level baraza combines both information and deliberation components, it also works through the (combined) theories of change of these components (Figures 1 and 2).

The broad nature of the baraza intervention means that many issues can come up during the meetings. This may make it hard to, in advance, determine where impact emerges. If many communities struggle with the same issue (e.g. absences of functioning toilets for girls at the public primary school), it will be easier to pick up an effect of the baraza on that particular issue, even though there may not be impact on the education sector as a whole. However, it may be that different communities struggle with different issues. In that case, it may be that no effects are found on particular issue, but all effects within a sector go in the same direction. Furthermore, primary outcomes are mediated by different channels, including enhanced contact with policy makers and service providers, increased citizen participation in elections, more cash and in-kind contributions to the commons, and changes in perceptions and prioritization. Some of these mediating channels are less specific and less localized.

- The Information Mechanism

Information treatments are only effective if a lack of information hinders the delivery of public service. We are confident that incomplete and asymmetric information is actually blocking effective service delivery here for three reasons. Firstly, information provision was one of the main aims of the baraza programme according to the OPM. Secondly, the hypothesis that information is key in our context derives from previous literature (e.g. Raffler, Posner, and Parkerson 2018). Thirdly, information frictions were named as a main constraint to public service delivery during our qualitative diagnostic work (Van Campenhout et al. 2018). We identified the baraza's potential to simply reduce information inefficiencies. For these reasons we confidently decided to isolate and explore the information component in the (existing) intervention.

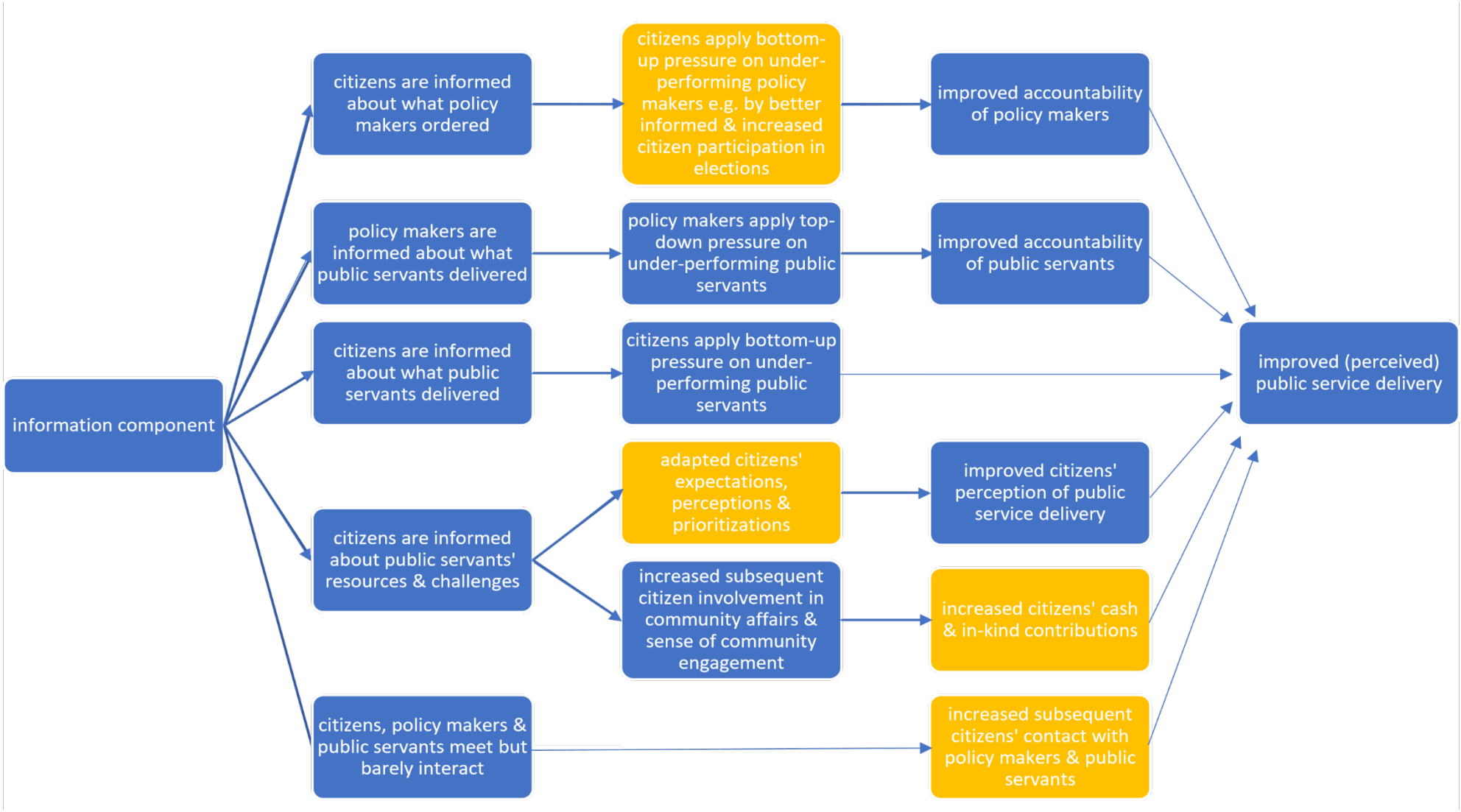

Figure 1 - Information Mechanism

In situations 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 reminiscent of the principle–agent problem. In essence, there are three players (elected politicians, civil servants and citizens) with only partly overlapping information sets and potentially competing interests[2]. Bringing stakeholders together in town hall type meetings is assumed to reduce information asymmetries. Increasing the knowledge of all stakeholders about what the client (policy makers) ordered and what the implementer (public servants) delivered may be an effective way to increase the quality of public service delivery by (1) allowing citizens to monitor and apply bottom-up pressure on under-performing civil servants and (2) increasing top-down pressure on under-performing civil servants by revealing to politicians the discrepancy between what was promised and actual performance, and therefore improving accountability of service providers. It can also improve the accountability of local public decision makers by allowing citizens to apply bottom-up pressure on under-performing policy makers, e.g. by participating more or better informed in elections.

There is some evidence that channelling 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).

The information component may also be important in managing expectations of the client. Citizens may have exaggerated beliefs about the resources at the disposal of decision makers and service providers, or they may not fully appreciate the challenges civil service providers face when doing their job. For instance, during focus group discussions, service providers mentioned that citizens sometime blame officials for things they have no control over. More in general, information may help sensitize citizens about the role of public service provision (for example, making sure boreholes are present) but that there are also limits to what citizens can expect (for instance, citizens are still required to boil water). When information can rectify inflated expectations and change perceptions, we may not find changes in the quantity or quality of public services, but we may still find changes in citizens' perceptions about the quality of these services. Informing citizens about the resources and challenges of public servants may also increase their involvement in community affairs and sense of community engagement, and therefore their willingness to contribute to the common goods. As citizens, policy makers and public servants meet during the information barazas, this might also raise subsequent contact of these stakeholders.

---

[2] As the public servant must be responsive to the needs of both the client and the community at the same time, the problem can be characterized as a multiple or common agency problem, which adds a collective action component to the standard principle-agent problem (Bernheim and Whinston 1986).

The Deliberation Mechanism

There are various ways in which deliberation increases 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 supports the positions articulated. This can change expectations, perceptions and prioritizations of citizens, and improve their perception of public service delivery. Secondly, deliberation can more effectively distil social choice than simple voting and majoritarian rule, in part by building of consensus both among citizens and between public servants and citizens. Policy makers and public servants are better informed about what citizens actually want. Thirdly, deliberation has been found to positively impact on the vigour and breadth of subsequent citizen involvement in community affairs (Björkman Nyqvist, de Walque, Svensson 2017). Deliberation provides opportunities for citizens to confront their leaders and public servants with issues and threaten with social and political sanctions if it is deemed that they are not performing. Citizens apply bottom-up pressure on under-performing policy makers, e.g. by better informed and increased citizen participation in elections, and on under-performing public servants, e.g. by social sanctions, and therefore improve public servants' and policy makers accountability. Creating a platform where stakeholders can meet and interact may also increase mutual understanding and create a better relationship between them. This can lead to more subsequent contact between citizens, policy makers and public servants. Elected official and service providers could also be more motivated because of this improved relationship. Furthermore, the mutual understanding and better relationship between stakeholders may also increase citizens' involvement in community affairs and sense of community engagement, and therefore their willingness to contribute to the common goods. However, when relationships are already poor, public fora that degenerate into name-and-shame sessions may make matters worse. Facilitated, collaborative meetings that jointly engage citizens and service providers in monitoring are often more effective than confrontational meetings (Waddington et al. 2019).

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 facts. If citizens can engage with policy makers and civil servants, they may request information that is relevant to them.

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

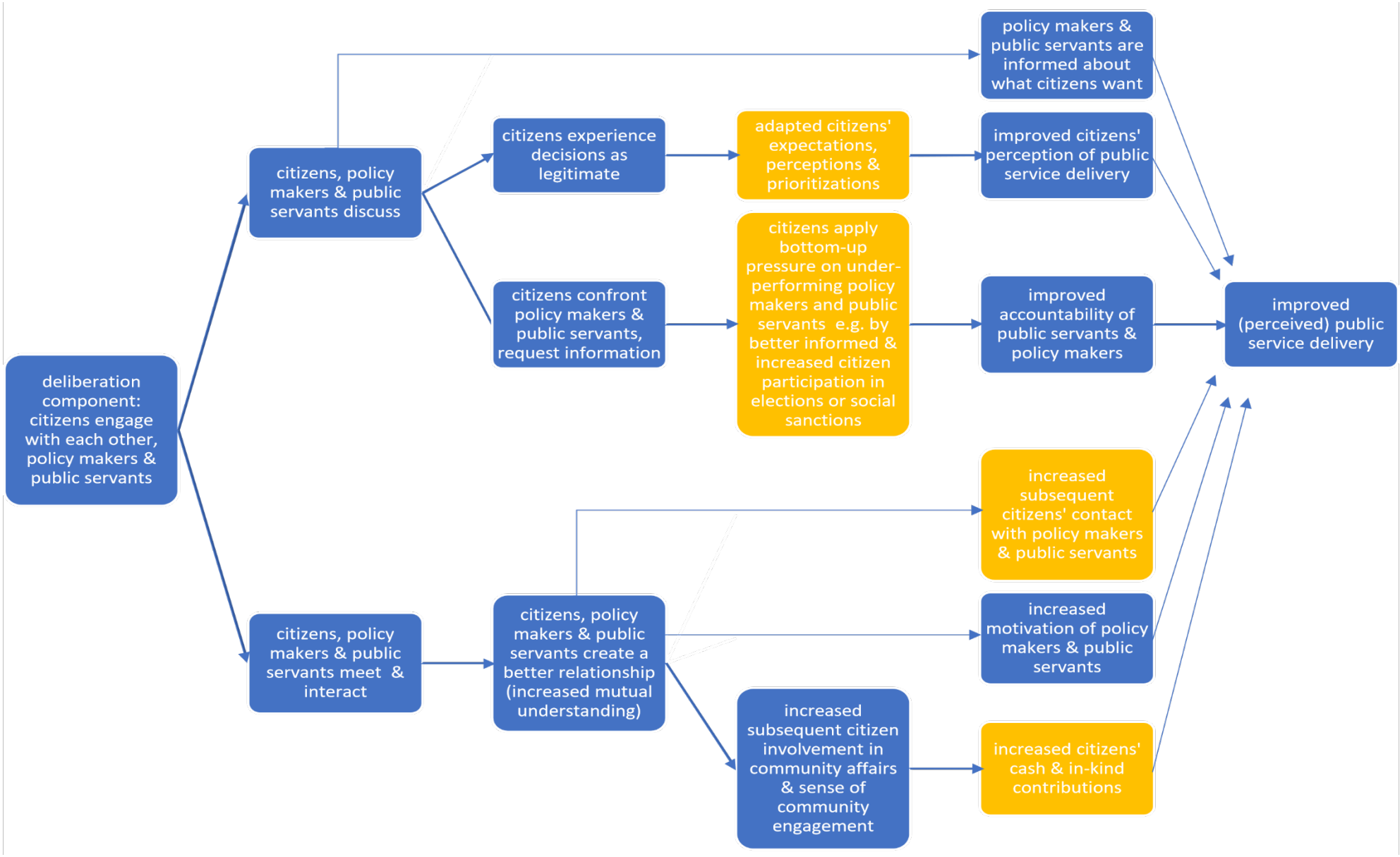

Figure 2 - Deliberation Mechanism

- 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 trade-off 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.

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, Duflo 2004). It may also be that issues highlighted at a local level fall under the responsibility of higher-level authorities and vice versa.

2.3 Intervention monitoring plan

After completion of the baseline, we trained local government officials and designated facilitators to ensure adherence to the intervention protocols. We agreed with OPM that for the barazas that were part of the study, facilitators would be selected from these trained facilitators. We developed detailed scripts that RDCs and facilitators were expected to follow. Furthermore, manuals for RDCs and facilitators were developed. Detailed information can be found in an online appendix.

Two full-time research assistants were also assigned to monitor program implementation. The research assistants worked very closely with the OPM staff tasked with the implementation of the barazas. One researcher accompanied OPM to all barazas that were part of the study. He also made sure that the information was disseminated in time for the information focussed barazas.

At the end of a baraza, the RDC is required to make 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. These reports were also collected to assess implementation and adherence to the intervention protocols ex-post.

3. Evaluation questions, methods, design, sampling and data collection

3.1 Primary and secondary evaluation questions

Primary evaluation questions look at the impact of the baraza programme and its key components on public service delivery. Four comparisons are made, corresponding to the following evaluation questions:

- What is the impact of sub-county level barazas on public service delivery in general, at sector level, and for selected individual outcomes that were preregistered?

- What is the relative importance of the information component of a sub-county level baraza on public service delivery in general, at sector level, and for selected individual outcomes that were preregistered?
- What is the relative importance of the deliberation component of a sub-county level baraza on public service delivery in general, at sector level, and for selected individual outcomes that were preregistered?
- What is the impact of district level barazas on public service delivery in general, at sector level, and for selected individual outcomes that were preregistered?

Secondary evaluation questions are looking into the mechanisms through which the baraza project is assumed to affect public service delivery. These include interfacing with politicians and civil servants, political participation, contributions to the common good. We also investigate how perceptions may have changed as a result of a baraza.

3.2 Methods

- Confirmatory analysis

In a first part of our analysis, we strictly follow a pre-registered analysis plan that takes the form of a “mock report”. This report was written in December 2019, just before end-line data was collected. It contains the results of an analysis on simulated end-line data for the four primary research questions. This mock report was pre-registered at the American Economic Association's RCT registry with a time stamp. Pre-registration and mock reports are effective tools against fishing and false-positive science (Humphreys, De la Sierra, Van der Windt 2013).

The mock report was prepared using Lyx, an open source Latex front-end. All Latex and R code to replicate the analysis was 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/]. The use of revision control further increases transparency and allows for easy replication (Ram 2013).

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

The report describes explicitly which variables will be used to assess impact, how they are combined into indices, and what transformations are used, referencing the actual names of the variables in the end-line data collection application.

This confirmatory analysis focusses on a subset of carefully selected and declared variables that are combined in indices – one overall index, and four indices corresponding to the sectors. Our indices combine individual outcome variables covering the use, availability, quality and delivery of services. These outcomes were categorized into four broad sectors: agriculture, health, education, and infrastructure (mainly drinking water and roads) following Anderson (2008) to account for multiple hypothesis testing. The four indices are then in turn combined into an overall indicator of public service delivery.

For continuous variables, 5 percent trimmed values were used (2.5 percent trimming at each side of the distribution). Inverse hyperbolic sine (IHS) transformations were used if skewness exceeded 1.96. Trimming was always done on end results. For example, if the outcome is yield at the plot level, then production was first divided by plot area, after which the inverse hyperbolic sine transformation is done, and the end result is trimmed. Outcomes for which 95 percent of observations have the same value within the relevant sample were omitted from the analysis to limit noise caused by variables with minimal variation.

- Detailed analysis

A second more exploratory part of our analysis looks at individual outcomes beyond the indices. Individual outcome variables cover a wide range related to the use, availability, delivery and quality of public service. Using ANCOVA models, some outcomes of each of the four key sectors will be compared between the different groups. Furthermore, we explore whether barazas affected aspects that are at the core of community-based monitoring, such as participation in elections of local leaders, citizens' contact with policy makers and service providers, perceptions of service quality and prioritization, and cash and in-kind contributions. Finally, we analyse sub-county level data that was collected from government officials because in addition to household surveys, we conducted surveys with 261 officials as respondents.

3.3 Evaluation design

- Study design and identification strategy

This study proposed a nested, or two-step, randomization design, illustrated in Figure 3. 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 (D^{ID}), while other districts do not receive a baraza at this level (D^0). 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 (S_0^0). About one quarter will receive a sub-county level baraza that combines both information and deliberation treatment (S_{ID}^0). A third quarter will receive a sub-county level baraza that consists largely of officials providing information and limited opportunity for citizens to engage (S_I^0). 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 (S_D^0). We also take a random sample of sub-counties from the D^{ID} districts that received the district level baraza (S_0^{ID}). Within each sub-county, we sample a fixed number of households.

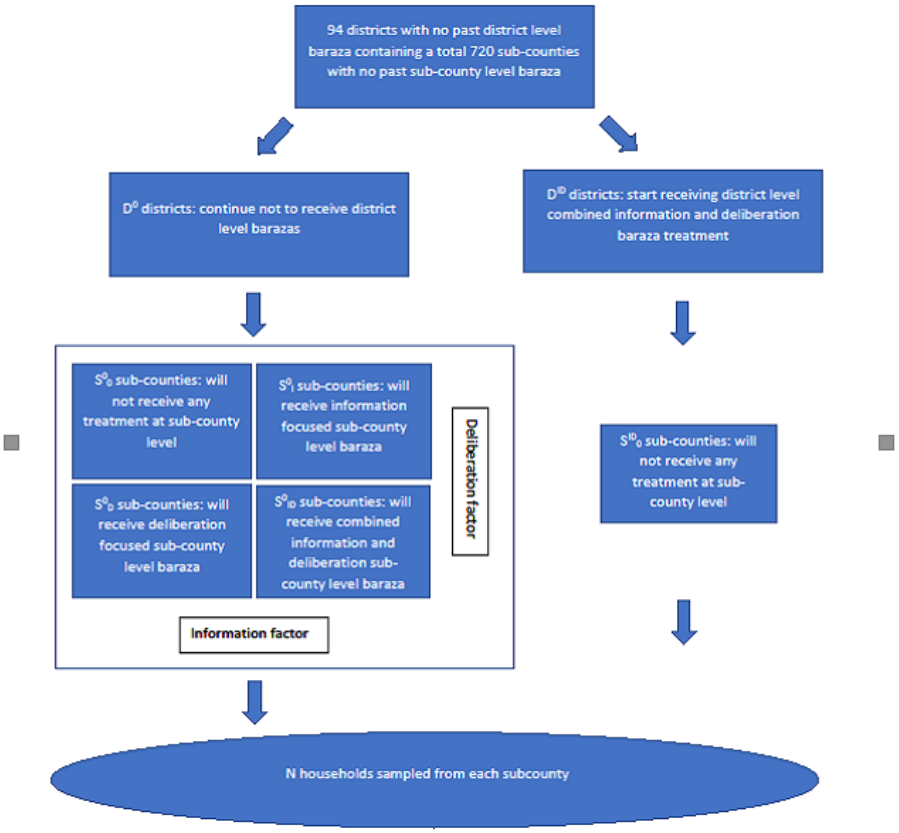

Figure 3 - Study Design

The above design allows us to answer the four research questions. 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 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 (S_0^0) or because they only received a deliberation focused baraza (S_D^0)). 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 (S_0^0) or because they only received an information baraza (S_I^0)). 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. Finally, to investigate administrative placement of the intervention, two comparisons will be used. First, and similar to the first hypothesis, we can simply estimate the impact of district level barazas by comparing outcomes of households that were sampled from S_0^{ID} to households that were sampled from S_0^0 . However, we can also directly compare district level barazas to sub-county level barazas by comparing outcomes of households that were sampled from S_0^{ID} to outcomes of household that were sampled from S_{ID}^0.

From a policy perspective, in light of the shift from sub-county level barazas to district level barazas from 2012 onward, the last comparison is the most interesting one, and this comparison was pre-registered. However, the partial roll-out of the intervention means that for this comparison we are constrained by the number of sub-counties in S_{ID}^{0} that ended up being treated. Comparisons of outcomes in areas that received a district level baraza treatment to areas that did not receive a baraza have more statistical power because more observations can be used.

- Power and sample size calculations

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 errors of the outcomes and inter-class correlations. Outcomes used to determine sample size included weight-for-age z-scores for children; number of days unable to work as percentage of days sick at the household level; number of years the average child within the household goes to school, proportion of children in the household currently attending school; the proportion of households that was visited by an extension worker in the previous year; maize yields; time to get drinking water (including waiting time); and share of households having access to improved drinking water sources. This resulted in the selection of a total sample size of 11,500 households distributed over 230 sub-counties 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.

The original power calculations assumed full roll-out of the intervention. However, due to implementation challenges that will be explained in detail in the section on intervention implementation fidelity below, a series of updated power calculations was performed prior to endline data collection. In particular, we simulated a new set of minimal detectable effects (MDEs) associated with the sample that we were about to collect[4]. We used baseline data to simulate MDEs for a selection of the outcomes we will use to judge effectiveness of the intervention (and are specified in the pre-registered report). We used a standard significance level of 0.05 (double sided).

Figure 4 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 during baseline data collection that they were visited by an extension officer in the year preceding the data collection. The grey solid line shows the power curve associated with the deliberation treatment, comparing the 1,900 households that received the deliberation treatment to the 3,450 households that did not receive a deliberation focused baraza[5]. The light blue dashed line closely tracks the grey 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 sub-counties that received an information baraza to the 2,900 households that did not receive a sub-county information baraza. The dark blue dashed line compares effectiveness of barazas

[3] We added an additional 3 sub-counties in each of the five treatment groups to account for attrition.
[4] Sample size was now largely determined by the extent to which OPM implemented the interventions.
[5] While sample size in treated areas was dictated by what was achieved by OPM, we did have some degrees of freedom in terms of the sample size in control areas. How the sample size in the control areas was determined is also explained in more detail in the section on intervention implementation fidelity.

conducted at different levels, with the MDE defined as the difference in outcome between 1,000 households that received the combined information and deliberation sub-county 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 that received the combined information and deliberation sub-county 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.

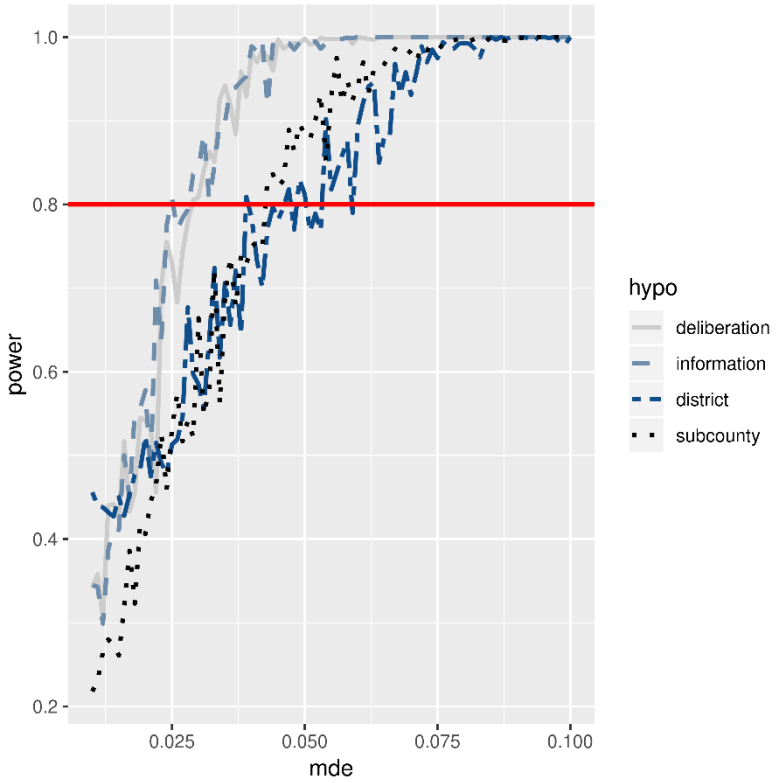

Figure 4 - Power curves for access to extension

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 percentage points. The deliberation experiment is similarly powered, and at 80 percent we can expect to identify effects of 3 percentage points or more. Due to the smaller sample size, comparing sub-county level barazas to pure control sub-counties seems harder. Here, the difference needs to be at least 4 percentage points. We have least power when comparing sub-counties level barazas directly to district level barazas, even though for this comparison, we have the same number of observations in the sub-groups than for the previous comparison. This is because the unit of randomization is at a higher level (districts rather than sub-counties).

In Figure 5, we plot MDEs for an infrastructure related outcome: distance in kilometre (km) to the primary water source during the dry season. We find that for the information treatment and the deliberation treatment, we can detect a 4 percent difference at the standard 80 percent power level. As the average household lives about 900 meters from the primary water source, this means we can identify effects in excess of 36 meters. Also here, the MDE is highest when directly comparing the effect of district level barazas to sub-county level barazas. Then, MDEs correspond to about 70-90 meters for the average household in our sample. On GitHub results are provided for similar power simulations for all the variables that will be use to judge impact of the baraza intervention in Section 4.2.2.

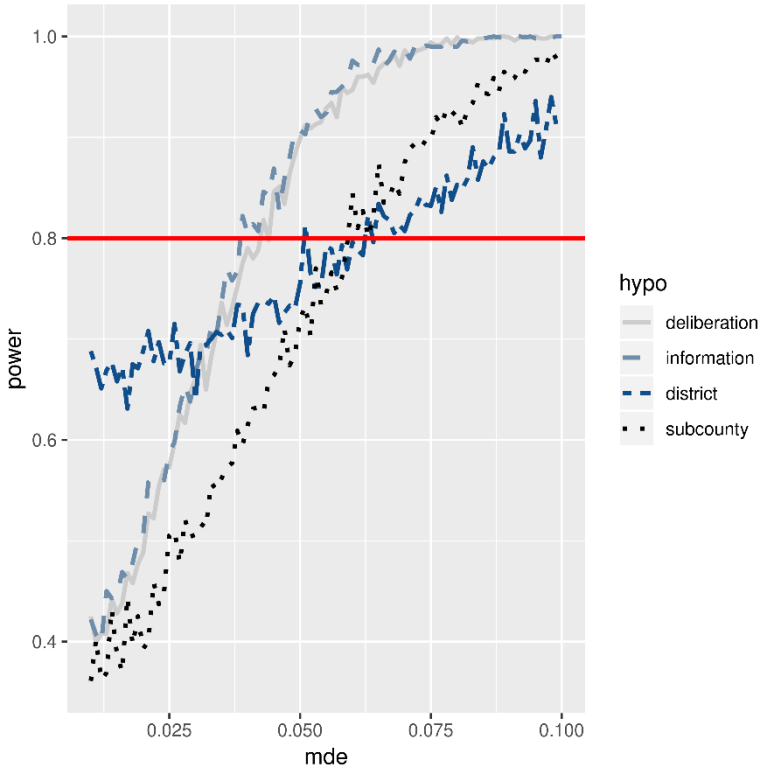

Figure 5 - Power curves for distance to water source

3.4 Ethics

This research was cleared by the Uganda National Council for Science and Technology (UNCST SS 5179), Makerere University School of Social Sciences Research Ethics Committee (MAKSS REC 05.19.291), as well as IPFRI's Institutional Review Board (DSG-19-1053).

During the implementation, we handled ethical protections carefully and followed protocols and procedures to minimize any potential harm in the field. Throughout baseline and end line interviews, we avoided asking sensitive questions such as those related to religion, political opinions or anything that was considered as too personal. Furthermore, we ensured that respondents could refuse to answer questions or abort the interview at any point and this would have no consequences. Interviewees would for example still get their token of gratitude if they refused to answer questions or stopped the interview. All field researchers are trained to assure privacy. In addition, we obtained sufficient permission from local authorities in the areas in which we worked. Also during data analysis we minimized potential harm. While we did collect identifiers to track respondents, data by which respondents could be identified or located were encrypted/ anonymized before being analysed or made public.

3.5 Sampling and data collection

We designed the experiment to cover 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 before, a small share of all sub-counties, albeit located throughout all of Uganda's 112 districts across the four regions, had already received a sub-county level baraza intervention. 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.

Figure 6 shows locations of households that were included in the study, clustered in sub-counties (blue, red, black and green) and in districts (orange). The colour codes denote the treatment to which the households were assigned. Blue denotes information only barazas, green deliberation only. Black are combined deliberation and information sub-county level barazas. Red are control sub-counties. Finally, the orange are sub-counties that are located in districts that received a district level baraza.

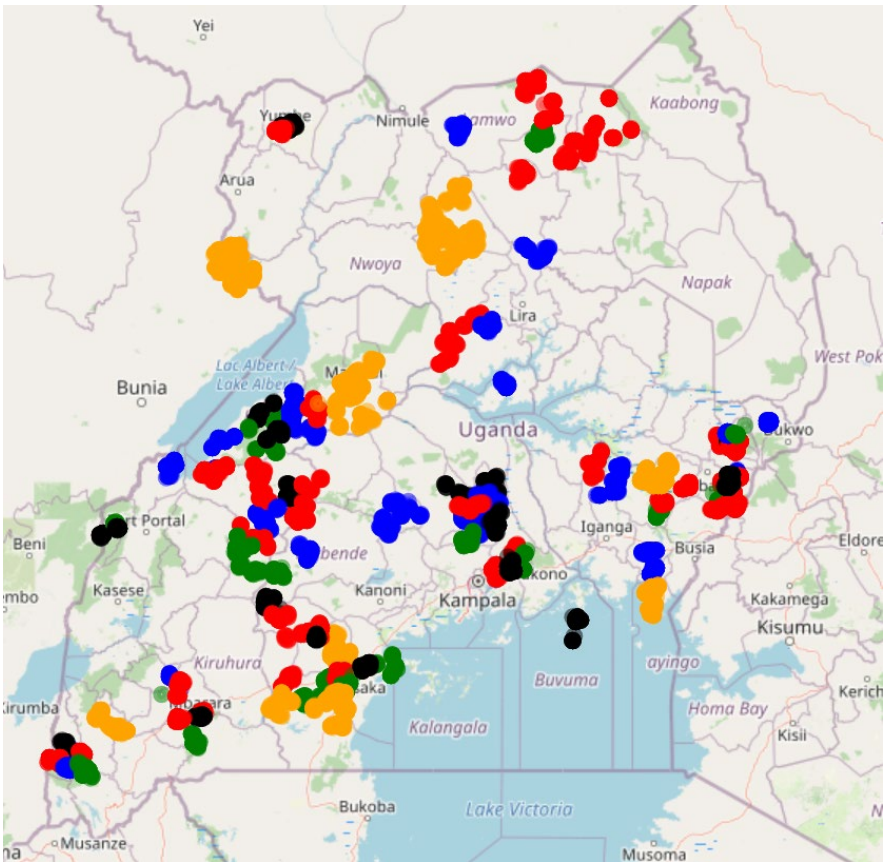

Figure 6 - Study area map

Data was collected using Open Data Kit. We developed and tested the tool in Bagezza sub-county in August 2019. We trained about 80 enumerators during a 3-day training in Kampala early January 2020 and rolled out the survey in the North, East and West simultaneously. Progress was tracked on a daily basis using Global Positioning System (GPS) mapping to trace out best routes and make sure areas were cleared.

The fact that the Open Data Kit (ODK) application already had many checks build in meant that little data cleaning was needed. Most of the code to run the analysis was ready, and as a result, a first report was ready by March 3rd.[6] Figure 7 provides a timeline.

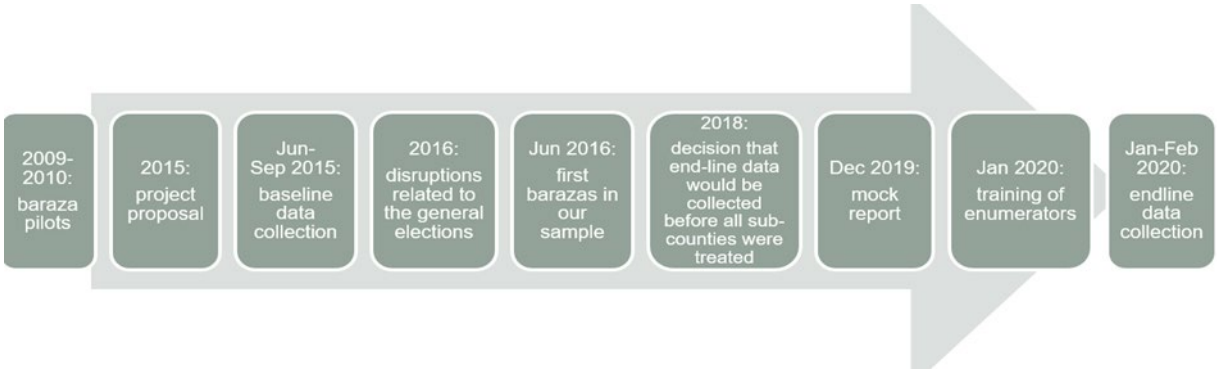

Figure 7 – Timeline

3.6 Intervention implementation fidelity

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 end-line data would be collected before all sub-counties 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, for logistical reasons, the implementing partner may have started with sub-counties that are close to the capital. Furthermore, OPM may have treated particular sub-counties first and other sub-counties later for political reasons. Our implementing partner may have selected politically preferred sub-counties first.

There are various ways in which we diagnose and remedy this potential problem. In the next section, we will present a series of balance tests. In particular, we will compare balance at baseline between subgroups as originally planned, and the final sample. We will also look at balance between households that were supposed to be treated but did not end up receiving treatment and households that were planned to be in the control group. This can be done for characteristics at both baseline and at end-line. We also propose to check robustness of the findings using a matching estimator.

In addition, 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). This raises the question: from the potential control sub-counties (either those that were allocated to the control

[6] This first version of the pre-registered report can be found here. However, later a coding mistake was found and corrected: For the comparison between full sub-county barazas to control barazas, we were reporting the interaction effect between information and deliberation instead of the combined information and deliberation effect. This was corrected on April 2nd in this commit (9f5afdbfdd6be766).

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 example, based on 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. This would mitigate the bias that would be introduced by a planner that prioritizes sub-counties in a particular location (for instance, close to Kampala).

We decided to use a range of sub-county characteristics that were likely to be known to OPM staff 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 CAOs 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 centres in the sub-county, female primary school dropout rate, number of Universal Primary Education (UPE) schools in the sub-county, percent of farmers that use improved seed, and political connections of the sub-county (defined by having a minister or member of parliament coming from the sub-county). These characteristics are used in a probit regression to predict the 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[7].

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

[7] A greedy matching procedure was use where we first calculate an adjacency matrix for all treatment and control subcounty populations. All these elements from the matrix where then ranked and those that were closest (in terms of the predicted likelihood of being treated) were selected.

Table 1 - Orthogonality tests for final sample

	mean	sub-county baraza	information	deliberation	district baraza
Household size	6.411	-0.186	0.065	-0.302	0.062
	(2.855)	(0.169)	(0.152)	(0.166)	(0.248)
Age of the household head (years)	47.009	1.096	-0.215	0.574	1.554
	(14.542)	(1.012)	(0.731)	(1.038)	(0.998)
Head of household is woman (1=yes)	0.191	0.025	-0.006	0.022	0.011
	(0.393)	(0.017)	(0.018)	(0.024)	(0.015)
Head finished primary education (1=yes)	0.208	0.005	-0.016	0.014	-0.018
	(0.406)	(0.029)	(0.025)	(0.035)	(0.031)
Thatched grass roof (1=yes)	0.262	0.015	0.044	-0.007	0.037
	(0.440)	(0.030)	(0.030)	(0.022)	(0.042)
Traditional mud wall (1=yes)	0.444	0.086	0.031	0.062	-0.008
	(0.497)	(0.058)	(0.053)	(0.058)	(0.114)
Distance to nearest all weather road (km)	0.909	-0.279+	0.027	-0.104	-0.229
	(0.912)	(0.136)	(0.140)	(0.135)	(0.112)
Access to extension (1=yes)	0.105	0.011	0.000	0.012	0.018
	(0.307)	(0.014)	(0.012)	(0.020)	(0.016)
Village Health Team in village (1=yes)	0.865	0.020	0.019	0.090*	0.075
	(0.342)	(0.051)	(0.036)	(0.039)	(0.041)
Number of children in public schools	2.507	-0.089	0.001	-0.188	0.078
	(2.072)	(0.118)	(0.097)	(0.111)	(0.154)
Number of observations	7,340	2,949	5,298	5,298	3,999

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports differences (and standard errors below) between baseline characteristics of households that received a district level combined information and deliberation baraza, and those that did not receive any baraza; **, * and + denotes significance at the 1, 5 and 10 percent levels.

While the ex-ante matching strategy may reduce bias resulting from incomplete roll-out, there are also costs involved. First, if sample selection is introduced by the roll-out, matching may further reduce external validity of the study, as now also the control sub-counties are not a random sample of the study population anymore. Second, the reduction in potential bias for hypotheses related to the sub-county level barazas should be traded off against an increase in potential bias when testing differences between control and district level barazas. As the sub-county level analysis weighed higher in terms of research objectives, we decided to prioritize the reduction of bias resulting from incomplete roll-out at this level. However, it should be kept in mind that both of these issues only become relevant if significant selection bias was introduced through the partial roll-out. Subsection 6.4 provides a more detailed discussion of the limitations of matching methods.

Figure 8 summarizes the factorial design that underlies the assessment of the relative effectiveness of the information and deliberation components of sub-county level barazas. As already noted in Section 3.2. 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 have followed the original power calculations and 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).

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

Figure 8 - Factorial design

However, the incomplete roll-out 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 8, 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 thus collected information on 49 sub-counties. As we wanted 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 end-line data (see next section), we made sure we selected half of these from the first column in Figure 8, and half from the second column. To test the impact of the deliberation treatment, we needed 38 control households. Also here, we made sure half were from the planned controls (first row in Figure 8) and half from sub-counties that were supposed to be treated, but were not (second row). Finally, as we also planned to directly test for the effect of a combined information and deliberation treatment, we needed at least 20 pure control sub-counties. Also here, we made sure half were selected from the upper left cell in Figure 8 and half from the sub-counties that were assigned to the treatment in the lower right cell of Figure 8 but did not get the treatment. Note that often, the same sub-county could 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 could be taken from the 14 sub-counties that were needed in that cell to test the impact of the information treatment. We thus simply took the higher number in each cell, which was 14 sub-counties. To allow for attrition, we selected 16 control sub-counties in each treatment cell.

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

Finally, we looked at the deliberation treatment. In this treatment, 38 sub-counties had been treated, either as deliberation alone or as part of the combined information and deliberation treatment, so we also needed 38 controls. We already had the 20 pure controls and an additional 11 controls from the previous steps. So we needed an additional 7 controls. As we wanted to investigate balance between control and planned but not treated controls, we selected 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.

4. Findings

4.1 Research analyses

In this section, we provide results for the four main hypotheses outlined in Section 3.1. Main results are presented in two parts. In a first part, we strictly follow a pre-registered analysis plan, and focus on a subset of carefully selected and declared variables that are combined in indices – one overall index, and four indices corresponding to the sectors. This confirmatory part of the paper will allow us to assess the overall impact of the baraza intervention.

A second part of the analysis is more exploratory in nature and looks at individual outcomes. In this part, we do not follow a pre-registered analysis plan. For each of the four key sectors – agriculture, infrastructure, health and education – a sets of outcomes will be compared between the different groups using ANCOVA models. We also explore if the baraza programme affected various aspects that are at the core of community-based monitoring, such as participation in election of local leaders, interfacing with politicians and civil servants, perceptions of service quality and prioritization, and contributions to public goods (both cash and in-kind). Finally, in this part we also report results for the analysis of sub-county level data that was collected from government officials.

4.1.1. Confirmatory analysis

Figure 9 provides a graphical representation of the overall impact on service delivery of the baraza, as well as on the different sectors. It shows the impact of the four main hypotheses–the impact of the sub-county baraza (sc baraza; indicated in grey), the relative effectiveness of the information component (info; light blue), the relative effectiveness of the deliberation component (delib; dark blue), and a comparison between sub-county and district level barazas (district; black) – on the four sectors we consider – agriculture, infrastructure, health, and education. The graphs are based on indices that are composed of individual outcomes in each sector as described in our pre-registered analysis plan, which are discussed in detail in the next section. We also combine the four indices into one overall index that assesses the impact on public service delivery in general.

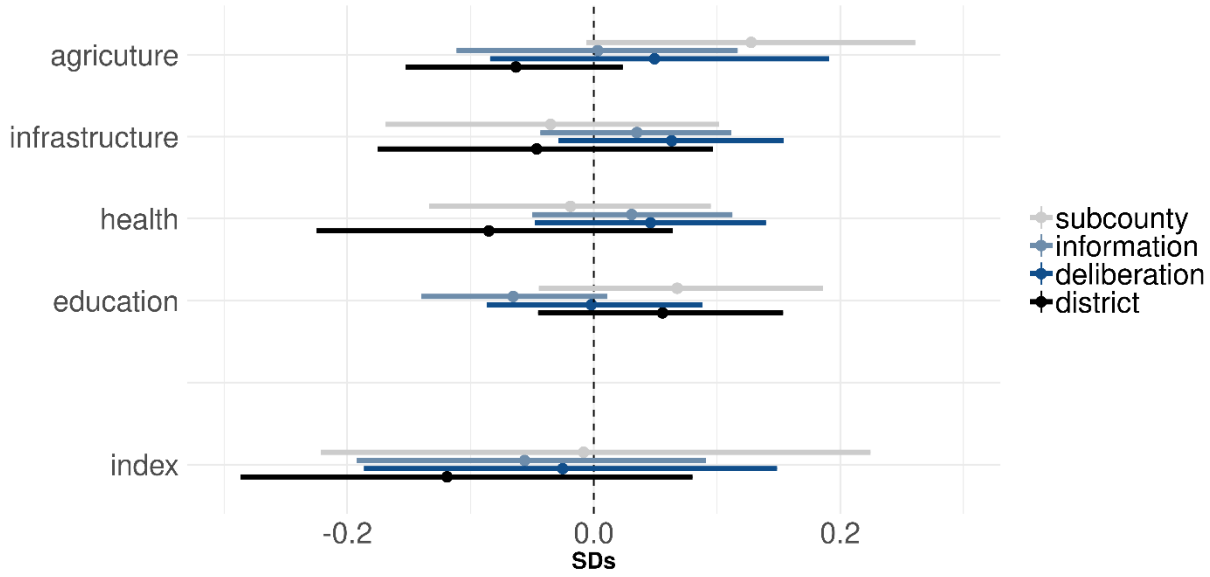

Figure 9 - Summary of baraza impact

The figure provides point estimates for the difference between treatment and control, estimated in an ANCOVA framework with controls for baseline outcome and region dummies. Confidence intervals are obtained following the permutation method explained in Gerber and Green (2012). This method first reconstructs a complete schedule of potential outcomes by adding and subtracting the average treatment effect for control and treated units respectively. These potential outcomes are then used to simulate all possible random allocations. For each allocation, average treatment effects are estimated and 2.5th and 97.5th percentiles are then taken as the lower and upper limits of the 95 percent confidence interval. This provides a conservative estimate of the confidence interval.

We find no significant impact of the baraza programme on overall public service delivery. There are some indications that sub-county level baraza did make a difference in the agricultural sector, but the difference is only significant at the 10 percent level. We do find that public service delivery in the agricultural sector was significantly worse in areas that were exposed to a district level barazas than in areas that were exposed to a sub-county level baraza.

4.1.2. Detailed analysis

The indices combine various outcomes, some of which the expected direction of the effect is unclear a-priori. For instance, an information baraza may increase the quality of services in a hospital or health centre when judged by an objective measure such as waiting time. However, the information may also result in higher expectations from the part of the user. As such, perceptions of quality may have reduced as a result of an information baraza. It is therefore also interesting to look beyond the indices and consider outcomes individually. This part of the analysis is more exploratory in nature and does not follow a pre-registered analysis plan.

- Agriculture

We first zoom in on the outcomes that are used to assess the effectiveness of barazas in changing service delivery in agriculture. Results are reported in table 2.

Table 2 - Impact of baraza on agricultural outcomes

	mean	sub-county baraza	information	deliberation	district baraza
Household used inorganic fertilizers?†	0.229	-0.015	0.034	0.001	-0.013
	(0.42)	(0.033)	(0.035)	(0.049)	(0.031)
Household used improved seed? †	0.364	0.043	-0.03	-0.037	-0.043
	(0.481)	(0.033)	(0.038)	(0.038)	(0.034)
Received improved seeds from govt?	0.121	0.051*	0.004	0.056	-0.005
	(0.326)	(0.024)	(0.025)	(0.043)	(0.015)
Household used agro-chemicals?	0.469	-0.028	-0.007	-0.005	-0.010
	(0.499)	(0.05)	(0.035)	(0.046)	(0.043)
Household used improved livestock inputs?	0.221	0.029	0.021	0.03	-0.014
	(0.415)	(0.031)	(0.028)	(0.034)	(0.026)
Did an agricultural expert visit your home? †	0.178	0.056**	0.037	0.036	-0.027
	(0.383)	(0.018)	(0.03)	(0.048)	(0.014)
Visited extension office/demo site/model farm? †	0.285	0.040	0.036	0.045	-0.013
	(0.452)	(0.028)	(0.035)	(0.044)	(0.023)
Are officials aware of extension demand?	0.264	-0.006	0.017	-0.001	-0.075*
	(0.441)	(0.024)	(0.027)	(0.035)	(0.023)
Not consulted for extension content?	0.316	0.034	-0.041	-0.031	-0.056
	(0.465)	(0.033)	(0.027)	(0.034)	(0.033)
Are farmer associations/groups in this village?	0.403	0.06+	0.04	0.087*	-0.032
	(0.491)	(0.03)	(0.038)	(0.041)	(0.027)
Farmer groups supported by govt? †	0.173	0.070*	-0.015	0.053	-0.037
	(0.378)	(0.028)	(0.03)	(0.04)	(0.022)
Received help in marketing from govt? †	0.069	0.018	-0.013	0.016	-0.014
	(0.254)	(0.022)	(0.017)	(0.017)	(0.012)
Received help in marketing from coop? †	0.062	0.037	-0.021	-0.001	-0.009
	(0.241)	(0.024)	(0.016)	(0.021)	(0.014)
Number of observations	6,704	2,738	4,858	4,858	3,687

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports differences (and standard errors below) between baseline characteristics of households that received a district level combined information and deliberation baraza, and those that did not receive any baraza; **, * and + denotes significance at the 1, 5 and 10 percent levels. † indicates that outcome was included in index.

We start by looking if the baraza programme affected the use of modern inputs in agriculture. A first outcome looks at whether the household used inorganic fertilizers (Diammonium phosphate (DAP), Urea, Nitrogen, Phosphorus and Potassium (NPK), Foliar, Triple Super Phosphate (TSP), Single superphosphate (SSP), Muriate of Potash (MOP)) during the last 12 months. The first column reports baseline averages, with standard deviation in brackets below. We find that about 23 percent of households in the sample give an affirmative answer to this question. In the second column, we report differences in outcomes between households that received a typical sub-county level baraza (in other words (i.e.) the crossed treatment of a sub-county information baraza and a sub-county deliberation baraza; the bottom right in Figure 8) and households that did not receive any baraza (pure control; the top left in Figure 8). We see that the proportion of households that reports using inorganic fertilizer is 1.5 percentage points lower among the sub-group of households that were exposed to a sub-county level baraza that consists of both the information and the deliberation component than among households that did not receive any baraza (second column). However, this difference is not statistically significant.

In the third column, we report differences between outcomes of households that live in areas where an information baraza was organized (either only an information baraza or a crossed

information and deliberation baraza; top and bottom right of Figure 8) 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 8). We see that adoption of inorganic fertilizer is 3.4 percentage points higher among households that were exposed to an information baraza. However, also here, the difference is not significant. In the fourth column, we report differences between outcomes of households that live in areas where a deliberation baraza was organized (either only a deliberation baraza or a crossed information and deliberation baraza; bottom left and right of Figure 8) and outcomes of households that live in areas that were not exposed to a deliberation baraza (either pure control or information only baraza; top left and right of Figure 8). We also do not find differences in terms of inorganic fertilizer use. Finally, in the fifth column, we compare households that were exposed to a district level baraza to pure control households. Again, no impact of the district level baraza is found on this outcome.

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

Next, we find that about 12 percent of households report that they received improved seed from the government extension system (that is through an extension agent, from the National Agricultural Advisory Services (NAADS) or through Operation Wealth Creation (OWC) that replace NAADS). We find that this is 5.1 percentage points higher in areas where a sub-county level baraza took place, and this difference is significant at the 5 percent level. We find no such effects from district level barazas. Direct comparison of district and sub-county level barazas indicate that sub-county level barazas are significantly more effective in increasing the likelihood that households reported to have received these inputs from government.

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

We then turn to advisory services. We first investigate if the barazas have affected access to extension at home. We estimate the percentage of households in our sample who report that they were visited by an expert (e.g. crop or livestock extension agent, or community-based facilitator or another experienced farmer) at the home in the last 12 months. We find that access to extension is low, with only about 18 percent of households reporting that they received such a visit. Interestingly we find that this percentage is significantly higher among households that were affected by a sub-county level baraza. The effect is large, amounting to a 30 percent

increase over the sample mean. Furthermore, the effect seems to come from a combination of the information and deliberation components; the components in itself do not seem to affect the outcome enough to render if significant. We also find that this effect is absent among households that live in sub-counties that received a the district level baraza. A direct comparison of extension at home between households that were exposed to a district level baraza and households that were exposed to a sub-county level baraza confirms that sub-county level barazas were significantly more effective. Comparing realized effects with MDEs indicate that we have sufficient power.

Home visits by extension officers are not the only way in which households have access to information. Extension offices, demonstration sites and model farmers are also an integral part of the Ugandan agricultural advisory system. Especially after the establishment of NAADS, such a demand-led service component that can be consulted by farmers when the need arises became more important than the more supply-driven component of training and visit. We thus also enquire if anyone in the household visited an extension office, demonstration site, or model farmer in the past year. We find that about 28 percent of households in our sample report access to extension in this modality. While the results are in line with extension visits at home, differences are not significant. Access to extension, both at home or though extension offices and demonstration sites, was also included in the agriculture index.

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

Related to the previous outcome, we ask how decisions related to what topics to cover in agricultural extension are made. We define this outcome in a negative way, that is, the indicator is true if decisions are made without consultation. We see that about 30 percent of households indicate that no consultation happens, and the content of extension advisory services is decided upon by experts at the central level. We do not find that the baraza intervention increased participation in extension service planning.

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

The final two questions focus more on marketing. Connecting farmers to markets is also an important strategy outlined in the Agriculture Sector Strategic Plan (ASSP). The first outcome relates to the likelihood that farmers are supported by government through the village procurement committee. In the sample, about 7 percent of households report that they were

assisted by government. A second questions is similar but looks at the role of cooperatives. We generally find no effects of the baraza, except perhaps for an increase of almost 4 percentage points in the likelihood that cooperatives assist with marketing in areas that received a sub-county baraza. Both of these outcomes were also included in the index to assess overall impact.

- Infrastructure

A second important area in which we expect to see an impact of the baraza programme is in infrastructure. We primarily focus on drinking water infrastructure. Results, similarly formatted as results in the previous section, are in table 3.

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

A second outcome we look at (and is also included in the index) is the distance to the primary water source during the dry season. This was measured in km, but trimmed and transformed using the inverse hyperbolic transformation. We find that, on average, households have to walk about 1 km. While this distance seemed to reduce in all comparisons, and especially for barazas held at the district level where we find a reduction of approximately 9 percent, it is never significantly different from zero.

The third outcome, also part of the index, is the time that one must wait at the water source, measured in minutes. This continuous variable was also trimmed and transformed. We find that households must wait on average about 37 minutes. We find a significant reduction in waiting time in areas that were exposed to the sub-county level baraza intervention, and some indication that the deliberation component is mostly responsible for this reduction.

Table 3 - Impact of baraza on infrastructure

	mean	sub-county baraza	infor-mation	delibe-ration	district baraza
Household uses unprotected water source†	0.159	0.031	0.005	0.010	-0.023
	(0.366)	(0.042)	(0.036)	(0.037)	(0.046)
Distance to water source (km) †	0.748	-0.026	-0.04	-0.049	-0.091
	(0.576)	(0.046)	(0.041)	(0.061)	(0.039)
Waiting time at source (min) †	3.198	-0.286+	-0.006	-0.287	-0.032
	(1.638)	(0.152)	(0.117)	(0.193)	(0.160)
Is there a Water User Committee in the village?†	0.598	-0.021	0.033	0.032	-0.009
	(0.49)	(0.046)	(0.037)	(0.04)	(0.047)
Is member of Water User Committee?	0.163	0.022	0.001	0.04	0.020
	(0.37)	(0.021)	(0.017)	(0.025)	(0.038)
Water User Committee holds public meetings?	0.474	-0.005	0.043	0.060	-0.050
	(0.499)	(0.044)	(0.036)	(0.042)	(0.056)
Satisfied with quality of drinking water?	0.624	0.031	-0.009	-0.062	0.002
	(0.484)	(0.052)	(0.044)	(0.044)	(0.019)
Treat water before drinking? (boil or treat)	0.5	-0.025	-0.087*	-0.02	0.010
	(0.5)	(0.045)	(0.037)	(0.046)	(0.041)
Distance to nearest all weather road (km) †	2.849	0.388	-0.129	-0.286	0.591
	(1.788)	(0.314)	(0.306)	(0.313)	(0.405)
Number of observations	6,704	2,738	4,858	4858	3,687

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports differences (and standard errors below) between baseline characteristics of households that received a district level combined information and deliberation baraza, and those that did not receive any baraza; **, * and + denotes significance at the 1, 5 and 10 percent levels. † indicates that outcome was included in index.

The fourth outcome variable assesses changes in the presence of a water user committee in the village. Overall, about 60 percent of households report that such a committee is present in their village. We do not find that this share varies between the different experimental groups. Similarly, we do not find that households are more or less likely to participate in such committees, nor that these committees hold more or less public meetings.

Households were also asked if they were satisfied with the quality of the water that is available at the source during the dry season. About 62 percent respond that they are satisfied or very satisfied with the drinking water. We do not find that households that are exposed to the baraza intervention are more or less likely to report that they are (very) satisfied with the quality of drinking water during the dry season. Half of the households report that they treat drinking water before drinking it, either by boiling it or treating it with chlorine. The likelihood that households treat water reduces somewhat for the information treatment. Potentially, better access the clean water reduces the necessity to treat water before drinking it.

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

- Health

We now look at outcomes in the health sphere (table 4). One problem with public health related outcomes is that some will only be available for households that have visited government health facilities, reducing sample size too much to maintain acceptable power.

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 his/her household had fever. We find that 70 percent of households respond that they would seek treatment in a government health facility. This proportion is independent of the treatment groups.

A similar indicator attempts to assess 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 the household was to give birth. This percentage is even higher than just for illness: more than 80 percent would go to a government health facility to give birth. Again, this proportion is not affected by the baraza programme. Both outcomes are included in the health index.

Next, we ask if a Village Health Team (VHT) is present in the village. VHTs are very important in front-line health care in Uganda. They also have prominent roles in government health interventions, such as immunization campaigns or the distribution of bed nets. We find that overall, nearly 90 percent of households report that a VHT is present in their village. The presence of a VHT is not impacted by the baraza intervention.

As the baraza tries to increase citizen engagement, we also check if households that were exposed to a baraza are more likely to participate in VHTs. We thus asked if any member of the household was a member of a VHT. We see that in about 10 percent of our sample, at least one household member is part of a VHT. The baraza intervention does not increase the likelihood that individuals participate as VHT members. Furthermore, the baraza intervention attempts to encourages sharing of information. As such, we expect that being exposed to a baraza may encourage VHTs to organize more public meetings. We find that overall, 43 percent of households state that VHTs have organized a public meeting in the last year. We find that this proportion is significantly higher in areas that were exposed to a sub-county level baraza. This effect seems driven by the deliberative component of a sub-county baraza.

We also consider distance to the nearest government health facility, measured in km. Overall, average distance to the nearest government health facility is almost 50 km. We do not find that barazas reduce this distance.

We then turn to health outcomes. We start by asking if any member of the household has been sick during the last year. This was the case in two thirds of the households in our sample. The intervention did not reduce morbidity in our sample. We then ask for each sick person in the household to record how many days he or she was ill, and use this to calculate the total number of sick days at the household level in the last year. The average household recorded almost 50 sick days according to this definition. We also do not find that the intervention affected the (trimmed and transformed) number of sick days. Finally, we look at the number of days household members were unable to go to school or to work, which provides an indication of the severity of illness. Calculated similarly to the previous outcome, we find that in the average

household about 35 school- or workdays are missed due to illness. Again, there is no significant reduction in this (trimmed and transformed) number. This last health outcome measure was included in the health index.

Table 4 - Impact of baraza on health sector

	mean	sub-county baraza	information	deliberation	district baraza
Seek treatment for fever in public health facility †	0.691	-0.008	-0.007	0.025	-0.010
	(0.462)	(0.033)	(0.033)	(0.040)	(0.046)
Go to public health facility to give birth†	0.813	-0.029	-0.033	-0.016	-0.070
	(0.390)	(0.034)	(0.029)	(0.035)	(0.043)
Is there a VHT in village? †	0.881	0.022	0.005	0.029	-0.019
	(0.323)	(0.031)	(0.025)	(0.025)	(0.027)
Member of VHT?	0.113	0.022	0.003	-0.001	-0.024
	(0.317)	(0.017)	(0.014)	(0.015)	(0.012)
VHT organizes any public meetings?	0.429	0.076+	-0.018	0.058	-0.046
	(0.495)	(0.041)	(0.033)	(0.040)	(0.036)
Distance to nearest govt health facility (km) †	3.875	0.256	-0.162	-0.252	-0.445
	(1.377)	(0.219)	(0.233)	(0.263)	(0.342)
Any members sick?	0.658	0.003	0.024	0.037	-0.015
	(0.475)	(0.023)	(0.028)	(0.033)	(0.024)
Number of days ill?	2.576	-0.005	-0.04	0.004	-0.064
	(2.189)	(0.091)	(0.149)	(0.166)	(0.105)
Number of days school/work missed due to illness†	2.273	-0.081	0.076	-0.006	-0.065
	(2.027)	(0.106)	(0.134)	(0.145)	(0.121)
Waiting time before being attended (min) †	4.744	-0.04	-0.133	-0.151	0.064
	(1.012)	(0.093)	(0.108)	(0.135)	(0.082)
Has visited traditional health practitioner? †	0.257	-0.017	0.016	0.034	-0.039
	(0.437)	(0.032)	(0.029)	(0.03)	(0.019)
Patient was examined by in-charge/doctor	0.411	0.044	-0.049	-0.070	-0.041
	(0.492)	(0.041)	(0.032)	(0.042)	(0.025)
Time of examination	3.403	0.048	-0.099	0.015	-0.002
	(0.761)	(0.066)	(0.070)	(0.091)	(0.083)
Paid anything	0.179	0.01	-0.008	-0.013	-0.005
	(0.384)	(0.023)	(0.025)	(0.042)	(0.024)
Received meds in hospital	0.709	0.000	-0.003	0.000	-0.024
	(0.454)	(0.036)	(0.023)	(0.033)	(0.034)
Satisfied with services at hospital	0.682	0.048	-0.026	-0.038	-0.011
	(0.466)	(0.033)	(0.031)	(0.038)	(0.026)
Number of observations	6,704	2,738	4,858	4,858	3,687

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports differences (and standard errors below) between baseline characteristics of households that received a district level combined information and deliberation baraza, and those that did not receive any baraza; **, * and + denotes significance at the 1, 5 and 10 percent levels. † indicates that outcome was included in index.

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

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

One problem that often crops up in the health sector is absenteeism. To assess this, we ask who examined the patient in the health center. Ideally this should be the doctor or in-charge. If this person is absent, patients are generally examined by nurses or lab technicians. We thus construct an indicator that is one if the household responds that the patient was investigated by the doctor or the in-charge, and zero otherwise. Only in 40 percent of the cases, a qualified person appears to do the examination. The baraza does not seem to lead to less absenteeism. We also look at the time that the examination takes. The average examination in our sample took about 22 minutes. There is no change related to the intervention.

Health care in Uganda is supposed to be free. However, corruption is widespread and often patients are required to make payments to receive care. We find that almost 20 percent of households report that payment was required the last time they visited a government health facility. There is no impact of the intervention. Related, users often complain about a lack of drugs in government health facilities. We asked if, during the last visit to a government health centre, drugs were received (indicating that drugs were available). We also asked if drugs had to be purchased from outside of the hospital (indicating that at least some drugs were missing). While 70 percent of households report that they received medicines in the health centre, almost all of them also mentioned that they had to also buy drugs outside of the hospital. For neither indicator, the intervention seems to make a difference.

We further probe for a subjective assessment of the overall quality of care at the health facility. Most households report that they are satisfied or very satisfied with services received at the government health facility. This seems to increase in areas where a sub-county level baraza took place and there is also a sizable difference in outcomes when comparing sub-county level baraza outcomes to district level baraza outcomes. However, none of the differences are significant.

We considered several other health related outcomes that feature prominently in other studies. One key outcome in Björkman and Svensson (2009) is immunization. However, we already find close to 100 percent immunization rates in our baseline data. Another outcome is child mortality. Child mortality rates at baseline were estimated at 38 per 1000 live births, which was deemed too low to include in the endline analysis. Raffler, Posner, and Parkerson (2018) find similar child mortality rates at baseline and speculate 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.

Education

Education outcomes to assess impact of the intervention suffers from a similar problem as the one encountered with health outcomes: not all households in the sample have children in

school, and so for many of the outcomes related to education, sample size becomes small. This also affects the indices. Results are presented in table 5.

Table 5 - Impact of baraza on education

	mean	sub-county baraza	information	deliberation	district baraza
Number of children in UPE or USE†	1.797	0.149	-0.168	-0.078	0.021
	(1.914)	(0.139)	(0.101)	(0.109)	(0.136)
Distance to public school (km) †	1.42	0.025	-0.047	-0.044	-0.002
	(0.763)	(0.057)	(0.067)	(0.071)	(0.042)
Has complete boundary fence? †	0.347	0.064	-0.061	-0.057	-0.008
	(0.476)	(0.048)	(0.046)	(0.049)	(0.045)
Has electricity?	0.338	0.165**	-0.04	-0.017	0.035
	(0.473)	(0.049)	(0.042)	(0.049)	(0.038)
Has water facility? †	0.703	0.106*	-0.023	0.026	0.073
	(0.457)	(0.041)	(0.048)	(0.05)	(0.050)
Has PTA?	0.945	-0.007	-0.029	0.000	0.000
	(0.227)	(0.014)	(0.019)	(0.028)	(0.012)
Has School Management Committee? †	0.915	0.008	-0.034	0.002	0.037
	(0.279)	(0.024)	(0.023)	(0.033)	(0.020)
Informed about SMC?†	0.882	0.021	-0.036	-0.042	0.009
	(0.323)	(0.023)	(0.024)	(0.032)	(0.019)
Inspectors visited schools?†	0.639	-0.004	-0.075+	-0.035	0.015
	(0.48)	(0.051)	(0.043)	(0.048)	(0.036)
Number of observations	6,704	2,738	4,858	4,858	3,687

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports differences (and standard errors below) between baseline characteristics of households that received a district level combined information and deliberation baraza, and those that did not receive any baraza; **, * and + denotes significance at the 1, 5 and 10 percent levels. † indicates that outcome was included in index.

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

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

We also look at school infrastructure. First, we ask households if the primary or secondary school attended by any of their children has a complete boundary fence. In the complete sample, it was reported that only about 35 percent of schools have such a fence[8]. We also ask if the school has electricity and if there is a water source available in the school. We find that

[8] The lack of a fence was a frequent complaint from parents during qualitative work.

overall, about 34 percent of schools have electricity and about 70 percent have a water source. We find that sub-county level baraza seem to improve school infrastructure. We considered many other infrastructure related outcomes, such as the number of classrooms and availability of functioning toilets for both girls and boys, but baseline data suggested there were generally no issues related to these outcomes.

We also look at how the school is managed, and how parents are involved. For instance, we look at whether the school has a Parent Teacher Association (PTA) and a School Management Committee (SMC). Almost all schools have a PTA. We further find that 91 percent of households state the primary or secondary school attended by any of their children has a SMC. However, not all households are informed about SMC meetings. The baraza intervention does not seem to affect how schools are managed, how parents can participate, or how information is shared. Finally, we ask households if an inspector had visited the school in the year before the survey. We find that about 64 percent of households indicate that schools were inspected. Surprisingly, this proportion reduces as a result of the information component of a baraza.

- Contact with policy makers and service providers

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

We find that about 43 percent of households in our sample have met with the (village) LC1 chair in the month before the endline data was collected. The baraza intervention did not affect the likelihood that citizens meet with the LC1 using this definition. About 20 percent of respondents report that they met with the sub-county chief in the last year. We do not find that the baraza intervention changed this likelihood.

Furthermore, we see that the information component of the baraza increases the likelihood that citizens interface with the head teacher or with members of the school management unit. We also see that the information component of the sub-county level baraza increases the likelihood of meetings with water committee members. Finally, and similar to meetings with sub-county chiefs, few citizens report meeting with health unit management committee members. But all coefficients on sub-county level interventions are positive. The index also shows that the largest effect on meetings is due to the information component, but the effect is not significant.

Table 6 - Impact of baraza on meetings

	mean	sub-county baraza	infor-mation	delibe-ration	district baraza
LC1 chairperson	0.426	0.001	0.030	0.035	-0.034
	(0.495)	(0.025)	(0.035)	(0.048)	(0.024)
Sub-county Chief	0.196	0.031	0.035	0.053	-0.040
	(0.397)	(0.02)	(0.035)	(0.052)	(0.020)
Head teacher/ SMC member	0.486	0.038	0.058*	0.048	0.019
	(0.5)	(0.028)	(0.028)	(0.037)	(0.024)
Health Unit Management committee member	0.155	0.040	0.020	0.061	-0.021
	(0.362)	(0.024)	(0.036)	(0.051)	(0.017)
Water Committee Member	0.382	-0.016	0.060+	0.044	0.012
	(0.486)	(0.040)	(0.034)	(0.051)	(0.037)
Contact index	0.000	0.037	0.089	0.107	-0.036
	(0.649)	(0.036)	(0.059)	(0.094)	(0.033)
Number of observations	6,704	2,738	4,858	4,858	3,687

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports differences (and standard errors below) between baseline characteristics of households that received a district level combined information and deliberation baraza, and those that did not receive any baraza; **, * and + denotes significance at the 1, 5 and 10 percent levels.

- Participation in elections

A second key aim of the baraza programme is to increase citizen empowerment. One way in which citizens can influence policy is through political participation. We thus expect that the baraza intervention will affect the likelihood that citizens participate in elections at various levels. We also ask if any of the household members hold any political or traditional position. Results are in table A.3 in the Appendix.esults show fairly high overall participation in elections at various levels, and no impact of barazas. About 30 percent of households report that at least one member holds a political or traditional position.

- Cash and in-kind contributions

The baraza programme also attempts to increase a sense of community engagement. One way in which citizens can participate is though contributing to common goods such as public infrastructure, education of health services. We differentiate between cash contributions and in-kind contributions.

Table 7 shows that about 32 percent of households indicate that they made in-kind contributions to public schools in their community in the last two years. Overall, most in-kind contributions are targeted towards drinking water facilities, and least in-kind contributions were going to a dam or irrigation facility, which is consistent with the difference in public nature of these two facilities. Cash contributions are distributed similarly, except for the fact that contributions to bridges and roads generally take the form of labour contributions.

Table 7 - Impact of baraza on Contributions

	mean	sub-county baraza	information	delibe-ration	district baraza
In-kind contributions to the school?	0.321	0.006	-0.085**	-0.019	-0.059
	(0.467)	(0.032)	(0.032)	(0.035)	(0.031)
In-kind contributions to the health centre?	0.126	0.011	-0.03	-0.031	-0.061+
	(0.332)	(0.023)	(0.025)	(0.021)	(0.023)
In-kind contributions to the road/ bridge?	0.384	0.025	-0.039	-0.011	-0.052
	(0.486)	(0.043)	(0.037)	(0.037)	(0.035)
In-kind contributions to the drinking water facility?	0.452	0.047	-0.01	0.059	-0.004
	(0.498)	(0.046)	(0.042)	(0.038)	(0.043)
In-kind contributions to the dam/irrigation facility?	0.093	0.022	-0.024	-0.028	-0.019
	(0.291)	(0.031)	(0.020)	(0.029)	(0.020)
In-kind contributions to any government structure?	0.233	0.04	-0.073*	0.012	-0.025
	(0.423)	(0.034)	(0.029)	(0.034)	(0.029)
In-kind Contribution Index	0.000	0.063	-0.107+	-0.016	-0.093
	(0.609)	(0.068)	(0.057)	(0.058)	(0.057)
Cash contributions to the school?	0.382	-0.005	0.053	0.021	0.101*
	(0.486)	(0.026)	(0.035)	(0.039)	(0.028)
Cash contributions to the health centre?	0.121	-0.023	0.053	0.051	-0.014
	(0.326)	(0.024)	(0.035)	(0.040)	(0.017)
Cash contributions to the road/ bridge?	0.097	-0.017	0.001	0.021	0.001
	(0.296)	(0.022)	(0.015)	(0.031)	(0.013)
Cash contributions to the drinking water facility?	0.37	-0.044	0.107*	0.057	0.094
	(0.483)	(0.034)	(0.043)	(0.048)	(0.043)
Cash contributions to the dam/irrigation facility?	0.04	0.001	0.008	0.001	-0.003
	(0.197)	(0.015)	(0.012)	(0.014)	(0.011)
Cash contributions to any government structure?	0.26	0.008	-0.027	0.007	0.029
	(0.439)	(0.030)	(0.026)	(0.039)	(0.030)
Cash Contribution Index	0.000	-0.033	0.076*	0.063	0.067
	(0.536)	(0.041)	(0.037)	(0.056)	(0.032)
Number of observations	6,704	2,738	4,858	4,858	3,687

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports differences (and standard errors below) between baseline characteristics of households that received a district level combined information and deliberation baraza, and those that did not receive any baraza; **, * and + denotes significance at the 1, 5 and 10 percent levels.

We find that the information component of the sub-county baraza reduces in-kind contributions but increases cash contributions. For in-kind contributions, these reductions in contributions are especially for schools and for government or community buildings. The increase in cash contributions as a result of the baraza intervention is especially for drinking water infrastructure.

Perceptions and prioritization

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

Table 8 shows that households tend to agree more that access to drinking water sources is a serious problem as a result of the information component of a sub-county baraza. Households that received a deliberation focussed sub-county level baraza are also more likely to agree that drinking water is usually dirty.

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

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

Respondents seem to perceive agricultural service delivery as the most problematic area. Averages on the likert scales are fairly high when asked if farmers agree extension officers visit rarely, and that there is a lack of transparency in how farmers are selected to receive inputs from government. We see that the issue of transparency reduces somewhat after a sub-county level baraza, but the effect is not significant. A perception index that combines all outcome indicates only a significant difference between sub-county level barazas and district level barazas.

- Sub-county level analysis

In addition to household surveys, we conducted surveys with 261 government officials as respondents. Like in the previous subsection, we analysed data on agriculture, infrastructure, health and education. Obviously, sample sizes are much smaller here, and so results should be interpreted with this caveat in mind. This full analysis on sub-county data can be found in Appendix A.4.

Table 8 - Impact of baraza on perceptions

	mean	sub-county baraza	information	delibe-ration	district baraza
Access to a drinking water source is a serious problem	5.151	0.048	0.606**	0.410 +	0.143
	(3.264)	(0.265)	(0.223)	(0.227)	(0.160)
Drinking water is usually dirty	4.428	0.072	0.057	0.442+	0.049
	(3.129)	(0.232)	(0.199)	(0.229)	(0.254)
Access to a government health centre or hospital is a problem	5.819	-0.193	0.365+	0.016	0.290
	(3.092)	(0.273)	(0.218)	(0.261)	(0.150)
Government health centres or hospitals do not have relevant medicines	6.495	-0.412+	-0.027	0.018	0.169
	(3.024)	(0.204)	(0.182)	(0.206)	(0.168)
Staff at government health centres or hospitals are rude to patients	5.040	-0.048	0.015	0.096	0.053
	(2.913)	(0.224)	(0.155)	(0.205)	(0.165)
Medical staff at government health centres or hospitals are often absent	4.776	0.032	0.081	0.127	0.301
	(2.757)	(0.173)	(0.142)	(0.202)	(0.152)
Access to a government primary school is a serious problem	4.930	0.032	0.046	0.021	0.037
	(2.905)	(0.246)	(0.205)	(0.21)	(0.227)
Teachers in government schools are often absent	4.847	-0.074	0.011	-0.061	0.124
	(2.72)	(0.182)	(0.17)	(0.211)	(0.189)
Children's learning outcomes in government schools are poor	6.360	-0.194	0.166	0.14	-0.246
	(2.918)	(0.18)	(0.155)	(0.187)	(0.154)
Availability/access to all-weather roads is a serious problem	5.157	-0.348	-0.023	-0.18	0.118
	(3.14)	(0.289)	(0.225)	(0.229)	(0.113)
Agricultural inputs supplied by the government are of poor quality	5.845	0.227	-0.027	-0.105	0.130
	(2.788)	(0.16)	(0.129)	(0.16)	(0.176)
There is lack of transparency in how farmers are selected to receive agricultural inputs from govt.	6.352	-0.351	0.22	0.042	-0.024
	(3.165)	(0.229)	(0.25)	(0.259)	(0.198)
Agricultural extension agents rarely visit.	6.372	-0.189	-0.001	0.103	0.007
	(3.218)	(0.268)	(0.301)	(0.344)	(0.233)
Agricultural extension agents are not aware needs of farmers.	6.098	-0.01	0.082	0.13	0.458+
	(3.074)	(0.224)	(0.254)	(0.321)	(0.162)
Perception Index	0.000	-0.033	0.035	0.026	0.031
	(0.514)	(0.039)	(0.033)	(0.037)	(0.030)
Number of observations	6,704	2,738	4,854	4,854	3,685

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports differences (and standard errors below) between baseline characteristics of households that received a district level combined information and deliberation baraza, and those that did not receive any baraza; **, * and + denotes significance at the 1, 5 and 10 percent levels.

4.2 Heterogeneity of impacts

- 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 end-line data collection. For instance, the first barazas were held around June 2016 (about one year after the baseline) and so more than 3 years will have passed between treatment administration and end-line data collection.

For the most recent barazas, there will only be a few months between treatment administration and end-line data collection. One may argue that sub-counties or districts that were treated early on have been exposed to the program much longer and hence one may expect larger effects on a range of outcomes for these sub-counties or districts than areas that only recently received treatment. Places might have also been treated with barazas first because they were politically favoured. It is furthermore possible that our implementing partner made larger effort at the beginning than at the end of the program. The effects of these early interventions may thus be systematically different. At the same time, for some outcomes, effects of the baraza intervention may dissipate (or even reverse) over time as enthusiasm fades, plans are abandoned and promises forgotten.

We find that the OPM organized quite a few barazas in May 2019. We thus reran the analysis and added an interaction term between the treatment indicator and an indicator variable that takes the value of one if the baraza that the household was exposed to happened more than 1 and a half years before the endline data collection (the indicator is coded as zero for the control group). Results are summarized in Figure 10 below. It displays average treatment effects for the four hypotheses on the four families of outcomes, and one overall index, similar to the summary in Figure 9.

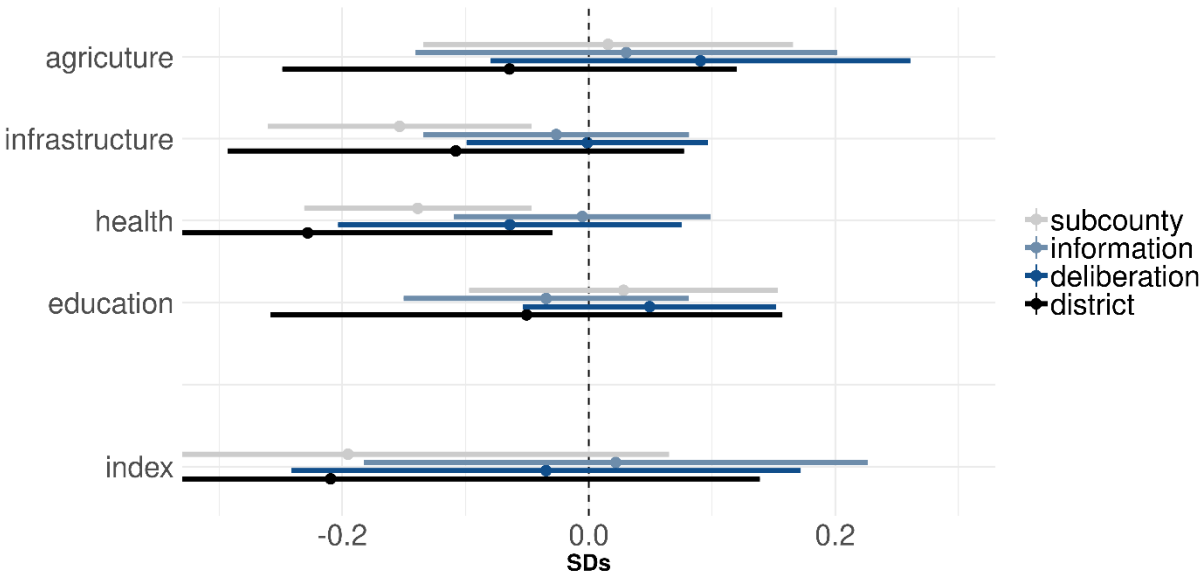

Figure 10 – Heterogeneity at SC level – effects more than one and a half years after implementation

For the agricultural sector, we do not find that the time elapsed between the intervention and end-line data collection affects the impact of sub-county level barazas we found in Figure 9. We find a significant negative interaction effect from sub-county level barazas for the infrastructure index. For outcomes in the health sector, we find negative interaction effects for sub-county level barazas and for district level barazas. The negative interaction effect for the sub-county level baraza seems to be driven by a reduction of the likelihood that households turn to government health facilities for maternal health in the long term. There are indications that these households are switching back to traditional healers, which may point to disappointment in the lack of progress made in public health facilities. For education, we do not find significant interaction effects based for the index. We do find a positive effect of sub-county level barazas

on enrolment in the longer run. Overall, results suggest that the effectiveness of barazas seem to dissipate over time.

- Heterogeneity related to officials recalling barazas

Our treatment indicator is based on information from the implementing partner. However, we also asked officials at sub-county headquarters if they recall if a baraza took place in the last five years. We also use this variable to check for heterogeneous treatment effects. Results are in Figure 11.

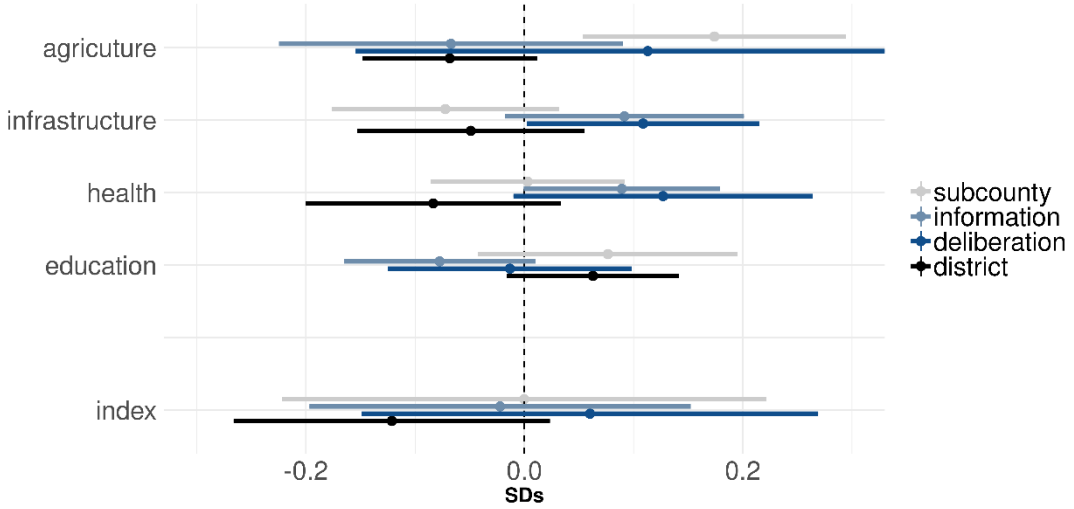

Figure 11 - Heterogeneity at SC level – officials recall baraza

Even though restricting to households that live in a sub-county in which officials recall that the baraza took place in treatment areas reduces sample size by only about 25 percent, we find substantial changes in the results. For the agricultural sector, differences with the full sample are minor. If anything, the positive effect of sub-county barazas on public service delivery in agriculture stands out even more. For infrastructure, the deliberation component seems to increase service provision, driven by a significant reduction in distance to water source.

Interestingly, we now also find significant results for the health sector. The deliberation component of a sub-county baraza increases the likelihood that households seek treatment in government health facilities when ill. The information component is associated with increased use of government health facilities for maternal health, and also increases the likelihood that a VHT is present. Both components also reduce waiting time before being attended to. There are no effects on education service delivery.

The generally larger impacts that we find in this sub-sample, particularly for the health sector, are intriguing. Potentially, officials that recall the baraza are intrinsically more motivated and thus more receptive to community-based monitoring. Alternatively, it may be that the information we received from the OPM is inaccurate and some sub-counties that they indicated as being treated were, in fact, not.

- Heterogeneity related to remoteness

Differences in the timing of the treatment and the fact that an official recalls the baraza introduces heterogeneity at the treatment level. However, heterogeneity may also depend on household characteristics. As outcomes are likely to be correlated within sub-counties, we will

have more statistical power to assess heterogeneity related to household characteristics than heterogeneity that originates at the treatment level.

One potential source of treatment heterogeneity at the household level is related to remoteness. Indeed, during discussions with stakeholders, it was often argued that barazas may have different effects on households that live close to the sub-county headquarters (HQ) versus those that live in remote areas. At baseline, we collected data on the distance between the homestead of the household and the sub-county headquarter. We find that this median distance is 5 km, and rerun the regressions but only for households that live 5 or more km away from the district headquarter. Results are summarized in Figure 12.

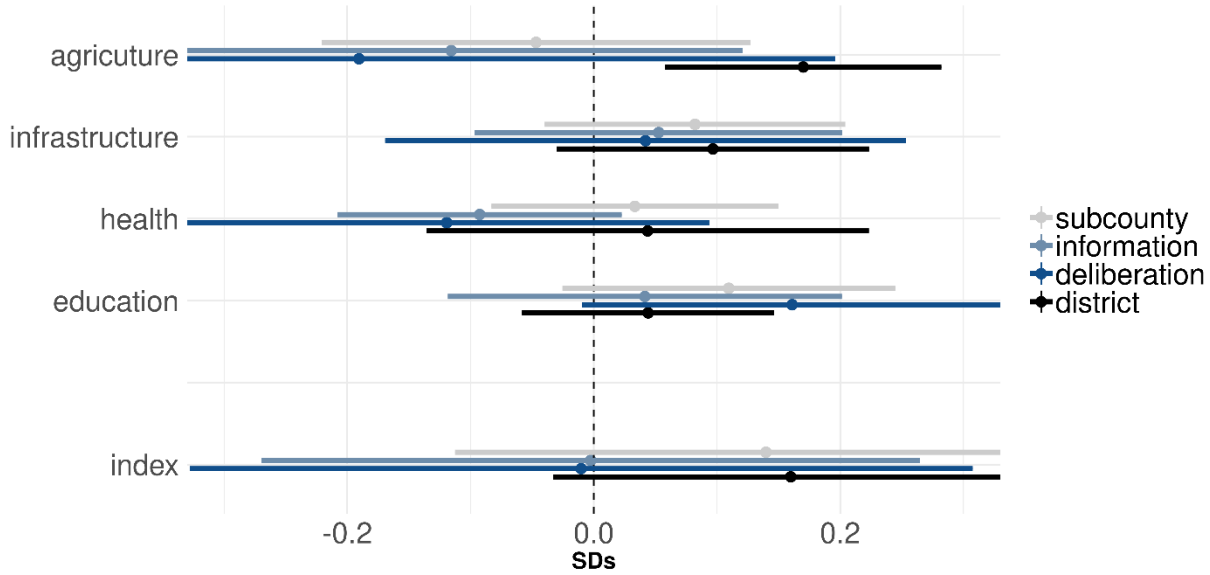

Figure 12 - Heterogeneity at individual level – living >5 km from SC HQ

There are some indications that district level barazas are effective for households that live further away from the sub-county. This is particularly the case for outcomes in the agricultural sector. For households in remote areas, we find that a district level baraza leads to a significant increase in the likelihood that households are visited by extension officers at home, the likelihood that they visit an extension office, demonstration plot or model farmers, the likelihood of being assisted by NAADS, and the likelihood of being supported by a marketing cooperative. Remote households also benefit from sub-county level barazas in terms of access to a protected water source, while district level barazas reduce waiting time at the source for this sub-group. We also find more positive treatment effects in the education sector if we focus on households that live further from the district headquarters. Overall, there are some indications that subcounty and district level barazas are particularly effective for households living in more remote locations.

- Heterogeneity related to households being aware of barazas

Out of the total sample of 6,700 households, we find that about 3,160 households respond that they are aware of the concept of baraza (and about 1,750 report that they remember that in the last 5 year such a meeting was held in their sub-county). Being aware of the concept of baraza may indicate that one is better informed or more interested in governance and public service delivery, which may also be an important source of heterogeneity at the individual level. We

thus reran the analysis, but only for the subset of households that indicated that they are aware of the concept of barazas. Results are in Figure 13.

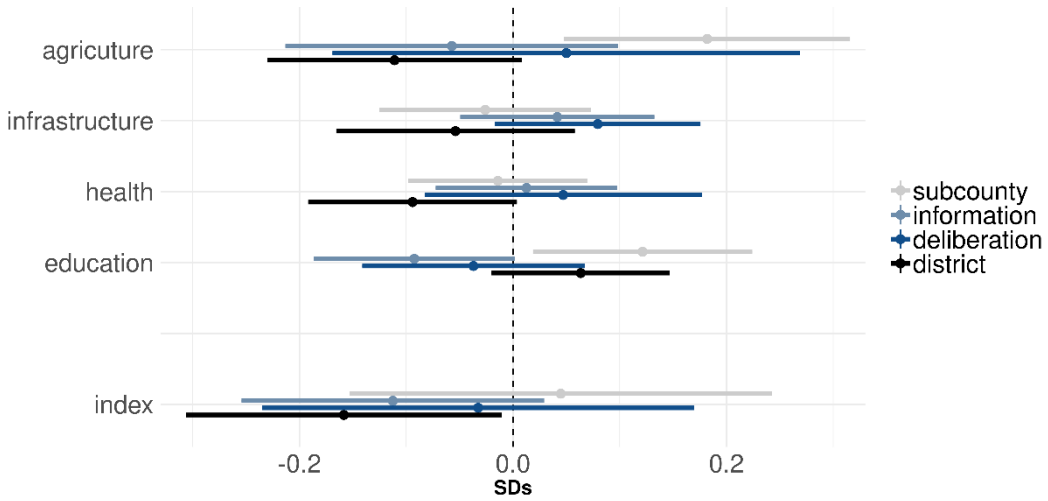

Figure 13 - Heterogeneity at individual level – knows baraza

We find that the positive impact of sub-county level barazas on agriculture becomes stronger. The effects are driven by more extension officers visiting and an increase in the support of NAADS/Operation Wealth Creation in the village. This suggests especially households that are well informed and interested in public service provision are able to cash in on the baraza. We also see a clear effect of sub-county level barazas on the education sector. While this effect is caused by increased enrolment in public schools, we also see significant positive effects on school infrastructure such as fencing and access to water on the school premises. The information component of the baraza significantly reduces outcomes. Potentially, providing only information (but no voice) to people that are receptive to participatory governance may lead to frustration, causing them to view some of the outcomes in a more negative light. This explanation is consistent with the results of the underlying individual outcomes. While outcomes such as enrolment rates and school infrastructure are not different between households exposed to information and those not, we do find that the former group complains significantly more about not being informed about the School Management Committee.

4.3 Threats to validity/Robustness

In this study, the primary treat to validity is the possibility that the partial roll-out introduced selection bias. We already showed in table 1 that the updated balance table that compares a range of baseline characteristics of actual treated households to control areas displays similar balance than the original balance table comparing planned treatment areas to planned control areas. While this is reassuring, in this section, we present additional balance checks to further explore if the roll out of the intervention was not random.

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.[9] 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

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

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 9 presents the original balance table (table A.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 (planned) control sub-counties 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 the 1 percent significance level once, at the 5 percent level once, and at 10 percent level once. Also here, this would be expected by pure chance alone, and so we conclude that the partial roll-out did not seem to have introduced selection bias.

We also revisit the risk that the partial roll-out may have introduced selection bias. While results in Table 1 and 9 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, we repeat the comparisons between control sub-counties and sub-counties that were allocated to a treatment cell but ended up not being treated from Table 9, but now use end-line data. Specifically for this reason, instead of simply collecting end-line data from the (planned) control sub-counties, we also collected end-line data from sub-counties that were supposed to receive a treatment but did not get one. Table 10 thus compares end-line 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 9.

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

We also find some imbalance when investigating the relative importance of the information component. Here, the imbalance is cause by the two variables that measure assistance in marketing. Also here, OPM may have prioritized where cooperatives and village marketing committees are less active[10].

[10] Note that in all cases, the bias is in the conservative direction, likely to lead to an underestimate of the treatment effect. We would be more worried if we found for example that households that were planned to receive a treatment but did not end up getting it had 10 percent lower incidence of visits by extension workers. This may indicate that OPM selected areas where extension was already stronger than in

Table 9 - Balance between planned but not treated sub-counties and planned controls

	mean	sub-county baraza	information	deliberation
Household size	6.324	0.012	0.388*	0.022
	(2.825)	(0.171)	(0.170)	(0.140)
Age of the household head (years)	46.501	0.357	0.698	0.553
	(14.615)	(0.714)	(0.663)	(0.808)
Head of household is woman (1=yes)	0.191	0.008	-0.019	-0.003
	(0.393)	(0.017)	(0.016)	(0.017)
Head finished primary education (1=yes)	0.213	-0.007	-0.007	-0.003
	(0.410)	(0.019)	(0.027)	(0.022)
Thatched grass roof (1=yes)	0.298	-0.002	0.000	-0.036
	(0.457)	(0.029)	(0.024)	(0.027)
Traditional mud wall (1=yes)	0.424	0.007	-0.057	0.044
	(0.494)	(0.049)	(0.047)	(0.044)
Distance to nearest all weather road (km)	0.906	0.284**	0.010	0.187
	(0.915)	(0.131)	(0.100)	(0.110)
Access to extension (1=yes)	0.108	0.005	0.008	0.007
	(0.310)	(0.015)	(0.016)	(0.015)
Village Health Team in village (1=yes)	0.854	-0.007	-0.01	-0.015
	(0.353)	(0.035)	(0.028)	(0.028)
Number of children in public schools	2.478	0.043	0.249+	0.076
	(2.074)	(0.112)	(0.115)	(0.100)
Number of observations	12,545	4,293	7,842	8,391

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

average areas, and this higher incidence of extension visits would erroneously be attributed to the baraza intervention.

Table 10 - Difference between planned but not treated sub-counties and planned controls at endline

	sub-county baraza	information	deliberation
Agriculture index	0.174**	0.113	0.045
	(0.057)	(0.057)	(0.057)
Infrastructure index	0.026	-0.031	-0.024
	(0.073)	(0.071)	(0.071)
Health index	0.026	-0.028	-0.012
	(0.047)	(0.039)	(0.043)
Education index	0.093	-0.002	0.116
	(0.057)	(0.046)	(0.045)
Public service delivery index	0.161	0.004	0.075
	(0.083)	(0.070)	(0.069)
Number of observations	1,637	2,356	2,808

Note: First column reports difference (and standard errors below) of the sub-county level baraza intervention; Column 2 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 3 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.

Our pre-analysis plan prescribed that if we find evidence of imbalance between planned but untreated sub-counties and planned control sub-counties using end-line information, we would try to recover unbiased impact estimates using a matched difference-in-difference estimator. However, we find baseline outcomes do not predict end-line outcomes very well. When autocorrelations are low, there are large improvements in power to be had from using ANCOVA instead of difference-in-differences (McKenzie 2012). Therefore, we deviate from our pre-analysis plan and use matching and estimate ANCOVA models on this pre-processed data.

For the matching, we use Mahalanobis distance with coarsened exact matching, an extremely powerful method of matching (Iacus, King, Porro 2012). We match on (baseline values of) 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. 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 GPS coordinates. For the comparison between sub-county level barazas and district level barazas, we did not match on GPS coordinates, as this resulted in too many observations that could not be matched. End-line data is then merged to the matched data-set, and standard ANCOVA models such as those used in the main analysis are estimated. Subsection 6.4 discusses the limitations of matching methods.

Figure 14 below provides a summary similar to Figure 9. We see that matching does not change the main conclusions. However, there are some differences between the matched and unmatched results when looking at individual outcomes. We provide detailed results similar to those in table 2 up to table 5 in Appendix table A.5 to table A.8, and provide a brief discussion of the most striking differences here.

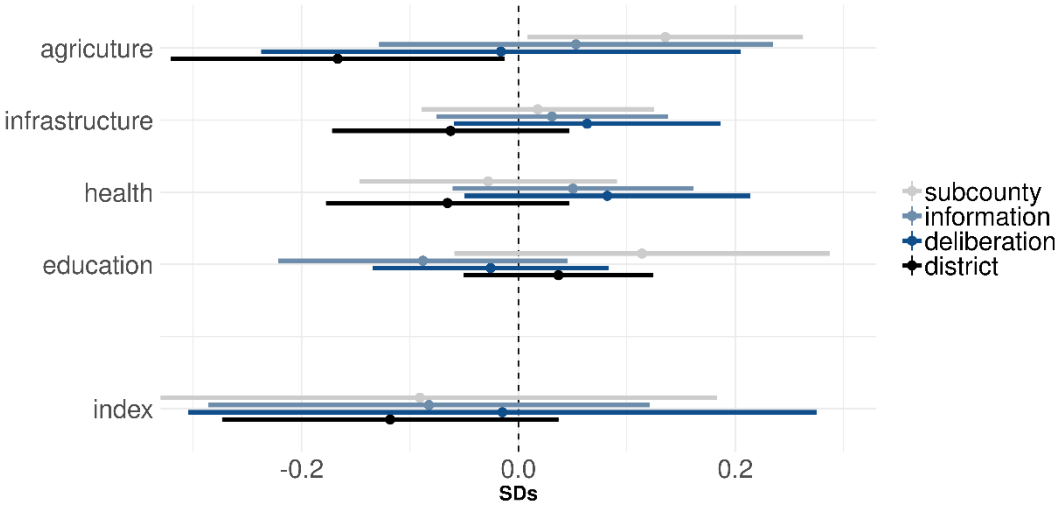

Figure 14 - Summary of baraza impact (matched ANCOVA)

For agriculture, comparing table 2 to Appendix table A.5, we see that after matching, the positive impact of the sub-county baraza on the likelihood of receiving seed from the government has disappeared. The positive effect of sub-county barazas on on-farm visits of extension workers is very similar. However, after matching, we also find a significant and positive effect of sub-county level barazas on the likelihood that a member of the household visited an extension office, demonstration field, or model farmer. This is consistent with the imbalance that we found above: that households in areas where a sub-county baraza was planned but not held reported a 10 percent higher incidence of visits to extension offices, demonstration fields or model farmers than farmers in areas where no sub-county level baraza was planned. We also find that it is more likely that farmer associations and cooperatives are present when a sub-county baraza was held. After matching, both information and deliberation components seem equally important in spurring the formation of cooperatives and associations. The analysis confirms that district level barazas are less effective than sub-county level baraza on a range of agriculture related outcomes.

For infrastructure, the reduction in waiting time at the water source as a result of sub-county barazas seizes to be significant after matching (Appendix table A.6). The negative impact of the information component on the likelihood that households treat drinking water also disappears. We now do find a positive and significant effect of the baraza intervention on citizen participation in water user committees. The effect seems to be driven by the deliberation component.

Further comparing table 4 to Appendix table A.6, we see that the effect of sub-county level barazas on the likelihood that village health teams organized public meetings persists. We also see that the somewhat puzzling negative effect of the information component on the likelihood that there is a functioning health management unit at the government health facility disappears after matching. This negative effect is replaced by a positive impact associated to a sub-county level baraza. Finally, we compare table 5 to Appendix table A.8 to assess the potential impact of non-random roll-out on the results for public service delivery in the education sector. Results are very similar. The negative effect of the information component on the likelihood that inspectors visit the school disappears after matching.

5. Cost analysis

5.1 Cost information

Cost information was provided by the OPM and is reported in table 11. OPM implemented 74 of 155 planned baraza forums. As such, the estimation provided here covers only the cost of the 74 barazas that were reported to have been implemented.

Table 11 - Baraza costs in absolute terms

	number of barazas	average cost (UGX)	total cost (UGX)	min cost (UGX)	max cost (UGX)
district baraza	7	15,325,000	107,275,000	13,900,000	18,500,000
sub-county baraza	20	12,837,500	256,750,000	11,200,000	14,000,000
information baraza	29	12,962,500	375,912,500	11,300,000	13,800000
deliberation baraza	18	12,712,500	228,825,000	11,100,000	14,200,000
total	74		968,762,500		

In total, implementing 74 barazas cost the Ugandan government about 968,762,500 Uganda shillings (UGX)[11]. A large share (39%) of this cost originated from sub-county information barazas. Due to the incomplete rollout and the distribution of the target sites from the center, and the resulting large differences in the number of barazas implemented for each type, the different types of barazas vary widely in total costs. However, the average cost of implementing different kinds of sub-county barazas does not differ a lot.

5.2 Cost effectiveness analysis

This analysis compares the costs of the different types of barazas (sub-county level baraza, information sub-county level baraza, deliberation sub-county level baraza, district level baraza) and their estimated effects. The costs were obtained from the OPM and are reported in table 11. However, assigning monetary values to the measures of effects is more challenging. Associating a monetary value to the effect of the baraza on the indices, for example to the effect of the sub-county level barazas on the agricultural index, is hard because the indices combine several individual variables. That is why we will focus on individual outcomes and assign monetary values to these individual outcome variables. We selected only individual outcomes for this cost effectiveness analysis, for which the estimated effect of the baraza intervention was significantly different from zero, as for non-significant outcomes, the estimated benefit of the intervention is zero. Among those significantly affected individual outcomes, we chose the ones for which monetizing is relatively feasible as it is for example impossible or at least inappropriate to attach a monetary value to the effect of barazas on the share of household that visits a public health facility to give birth.

A key objective of this study is to compare the effectiveness of barazas organized at lower administrative levels (the sub-county) to the effectiveness of barazas organized at a more aggregate level (the district). The level of administrative placement is an important determinant of the cost-effectiveness of the policy intervention: implementing a district level baraza affects far more people than implementing a sub-county level baraza, yet a district baraza costs only a little more than a sub-county baraza (table 11). As such, organizing a district level baraza could be more cost effective, even though the sub-county level baraza seems to have a larger impact at first sight.

[11] One United States (US) dollar corresponds to 3800 UGX.

For instance, in the area of infrastructure, we find that households have to wait on average about 37 minutes. A baraza intervention at the sub-county level reduces this time by about 29 percent, which corresponds to a reduction of about 11 minutes per household. Assuming a member of the household visits the water source once a day, the intervention saves 3,862 minutes (64 hours) per year per household. On average, 5,100 households live in one sub-county (Uganda Bureau of Statistics 2017, pp. 37-38), such that this sub-county intervention saves 19,698,393 minutes (328,306 hours) per year per sub-county. For a district level baraza, the impact is -3.2 percent, corresponding to 1.18 minutes every time a member of the household goes to the water source, totalling 432.16 minutes (7,20 hours) saved per household per year. However, an average 60,840 households live in one district (Uganda Bureau of Statistics 2017, pp. 37-38), such that this district intervention saves 26,292,614 minutes (438,210 hours) per year. The district intervention thus saved 109,904 hours more than the sub-county intervention. To attach a monetary value to this time difference, we consider the average hourly wage rate of 750 UGX. This results in a difference in impact of 82,428,180 UGX, while the district level baraza was only 2,487,500 UGX more expensive.

We now consider an example from the agricultural sector, namely access to extension visits at home. About 17.8 percent of households report that they were visited by an expert. A sub-county baraza increases this by about 5.6 percentage points, such that about 23.4 percent have access to extension at home, which corresponds to 1,193 households. However, the intervention at the district level decreases the access to extension at home by 2.7 percentage points. This means that the sub-county level intervention is more effective, as long as its benefits outweigh its costs. The 5.6 percentage points increase corresponds to 286 more households having access. From our baseline data we know that the average household farms 5.8 acres of land, which means that 1,656 acres are affected by the intervention. We use maize in our calculation because maize is an important crop in Uganda, both for home consumption and as a traded commodity because of its relatively high value-to-weight ratio. Average maize yields are at about 618 kg per acre for the main growing season, according to the Uganda National Household Survey 2005/06. Assuming that access to extension raises yields by 10 percent (Van Campenhout, Spielman, Lecoutere 2020), this results in 102,370 kg more maize produced due to the sub-county intervention. Assuming a bag of 100-kg of maize is sold at a median price of UGX 60,000, the monetary benefit of a single sub-county baraza, only considering access to extension, amounts to 61,422,278 UGX, while its average cost is 12,837,500 UGX.

For education, we consider the number of children in public schools to be an important outcome. As we can only find an effect on public service delivery in the education sector after deleting observations from households in sub-counties or districts where a baraza was held recently, we use this part of the analysis to compare the cost-effectiveness between sub-county and district barazas. We find that households have on average 1.79 children in school. A baraza intervention at the sub-county level increases this number by about 0.37 children per household. On average, 5,100 households live in one sub-county, such that this sub-county intervention leads to an additional 1,887 children in school. However, the intervention at the district level decreases this number by about 0.02 children per household. This means that the sub-county level intervention is more effective, as long as its benefits outweigh its costs.

Also for the health sector, we only find an effect on public service delivery after deleting observations from households in sub-counties or districts where a baraza was held recently. So again we use this part of the analysis. We look at the waiting time before being attended. The previous three examples it depends on the outcome whether a subcounty baraza is superior to

a district baraza or not. We now also compare the cost effectiveness of the information and deliberation components. Both reduce waiting time. Households must wait for 90 minutes, on average. The information treatment reduces this time by about 21.9 percent, which corresponds to a reduction of about 20 minutes per household, every time a member of the household visits a public health facility. Looking at our baseline data, we see that a member of the household visits this kind of facility 6 times a year, such that the information intervention saves 118 minutes per year per household. On average, 5,100 households live in one sub-county such that this information intervention saves 603,200 minutes (10,053 hours) per year per sub-county. For a deliberation baraza, the impact is -23.55 percent, corresponding to 21 minutes every time a member of the household goes to a public health facility, so 127 minutes saved per household per year. As 5,100 households live in one sub-county, this deliberation intervention saves 648,620 minutes (10,810 hours) per year. The deliberation intervention thus saved 757 (4,590) hours more than the information intervention. To attach a monetary value to this time difference, we consider the average hourly wage rate again. This results in a difference of 567,750 UGX, while both types of baraza are similar in costs (12,962,500 UCX vs. 12,712,500 UGX).

6. Discussion

6.1 Introduction

While we do not find that the baraza impacts public service delivery in general, we do find a variety of interesting effects when we look at individual outcomes and consider heterogeneity in the treatment effects. In light of this, our results confirm some of the likely explanations why Raffler, Posner, and Parkerson (2018) fail to find significant results on health outcomes in their study. For instance, we find indications that it may take some time for effects to materialize. The endline data in Raffler, Posner, and Parkerson (2018) was collected after 20 months, hence their results may only apply in the short run. In addition, the fact that our results are somewhat more encouraging than those found in Raffler, Posner, and Parkerson (2018) may also be related to the fact that our intervention is organized by the government. Raffler, Posner, and Parkerson (2018) find indications that presence of sub-county officials during the programming boosted the impact of the intervention on treatment quality in health centers. In line with this, community-based monitoring interventions organized by government may be more effective. This confirms that top-down monitoring may be more important in changing behaviour of civil servants than bottom-up monitoring by citizens.

This study focusses most on the analysis of the endline data of the quantitative component of the impact evaluation. A previous study also provides a less ambitious qualitative exploration of the likely impact of the baraza (Van Campenhout et al. 2018). In that study, we find that stakeholders think barazas are useful at improving public service delivery across all sectors, especially if the barazas took place at the sub-county level. Stakeholders had no difficulty providing examples of changes they felt were the direct result of the baraza being held: projects that were previously dragging were finished or taken up afresh; sub-standard work was redone; and in some instances, priorities were changed to better align with citizens' needs. A substantial part of these outcomes seemed to derive from the baraza's potential to fix information asymmetries. Focus group discussions suggested civil servants responded to the consequences of the increased likelihood of sub-standard work being exposed, and politicians responded to electoral considerations, suggesting barazas increased bottom-up pressure. There were also indications that barazas increased community involvement, as well as top-down monitoring.

The diverging results from the qualitative and quantitative analysis may be due to the fact that the baraza intervention is a broad intervention that attempts to address a range of issues in a heterogenous setting. It may be that the baraza is effective for some, but not for others. However, if a simple average treatment effect is estimated, the effect may turn out insignificant because it averages over subgroups. For instance, access to water is likely to be more of a problem in remote areas. Even if a baraza increases access to water and reduces waiting times, this may not show up if there is a large group close to the sub-county center that already has access to water and no additional boreholes were constructed in these areas.

This is illustrated when we link the endline data back to what we learned in the qualitative fieldwork (Van Campenhout et al. 2018). In Bagezza subcounty in Mubende District, drinking water was mentioned as a serious problem, and it was discussed extensively during the baraza. When we went back to the sub-county to test the endline tool, it appeared that the government made whole on their promises and the sub-county now had access to drinking water. To check this, we used baseline and enline data and simply compared means between Bagezza and a random control sub-county in the neighbourhood (Bwanswa in Kibaale district). Results are presented in the Figure below (Figure 15).

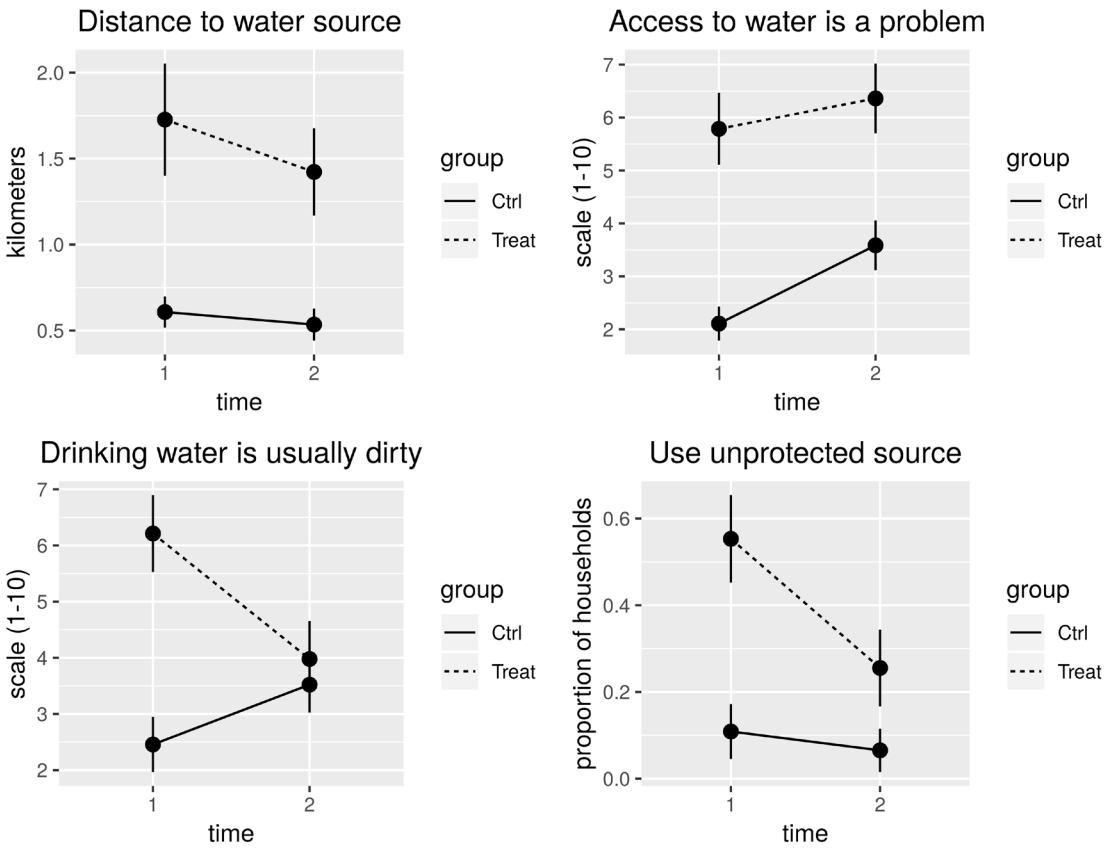

Figure 15 - Access to water in two sub-counties

The figure shows that on all four of the water related outcomes, Bagezza scored significantly worse when compared to the control sub-county. For instance, in Bagezza, average distance to the nearest water source was 1.75 km, while this only about 650 meter in the Bwanswa. The average score on the likert scale used to measure perceptions on the cleanness of drinking water is more than 6 in Bagezza, while this is only 2 in the Bwanswa. In Bagezza, more than half of all households rely on unprotected drinking water sources. In Bwanswa, this is only about 10 percent.

More importantly, we see that the difference between Bagezza and Bwanswa has reduced after the baraza happened in Bagazza. While still significantly higher than in Bwanswa, distance to the water source had reduced to about 1.35 km. Perceptions on access to water became more negative as a result of a general drought in East Africa, but less in Bagezza than in Bwanswa. The most impressive progress was made in terms of the quality of water. At the time of the endline, there is no difference between Bagezza and Bwanswa anymore. Use of protected water source increased over time in both sub-counties, but most dramatically in Bagezza.

Results from the section on heterogenous impacts are consistent with this explanation. For instance, we find that distance to water source is affected by the baraza intervention, but only if we restrict the sample to households that live 5 or more km from the sub-county district headquarters. Households living close to the headquarters may already have good access to water, and so a baraza may not affect their situation. Not taking this into account may lead to the conclusion that barazas do not influence access to water.

6.2 Policy and programme relevance: evidence uptake and use

At the time of writing of this study, we already presented preliminary findings at the National Monitoring and Evaluation Technical Working Group Workshop on march 13th at OPM, Kampala, Uganda. The response was very encouraging and OPM was very happy with the work so far. The meeting was attended by 80 persons, with representative from OPM, Kampala City Authority, Economic Policy Research Center of Makerere University, National Planning Authority, SDG (Sustainable Development Goals) Secretariat, Ministry of Finance, Planning and Economic Development, Ministry of Gender, Labour and Social Development, and about 10 other government authorities/departments. It is clearly too early to know how the evidence generated will be used, but we do feel that the study is held in high esteem by OPM.

6.3 Challenges and lessons

The mixed results are puzzling, especially given the fact that qualitative research prior to endline data collection suggested real effects from the intervention (Van Campenhout et al. 2018). We suspect that the lack of impact in the quantitative part of the study may be due to the nature of the intervention. Different sub-counties face different challenges, which is reflected in what transpires at the baraza event. For instance, in districts where there are issues related to water, the baraza will mainly revolve around poor service delivery in the infrastructure sector and how this can be improved. In these sub-counties, barazas may impact service delivery in infrastructure, but leave outcomes in other sectors unaffected. In other sub-counties, problems may concentrate in the agricultural sector, and impact on infrastructure may be minimal. In other words, the true treatment received by subjects may become hard to discern and may in fact be far from the standardized treatments given in RCTs in biophysical scientists. As a result, a focus on the average treatment effect may fail to identify a significant effect, as the impact is averaged over many sub-counties that in reality received a “different” type of baraza. Heterogeneity in the treatment will also introduce selection bias, as barazas will tend to focus most on areas that are most problematic (Barrett and Carter 2010). Issues related to non-standardized treatments are confirmed when looking at heterogeneous treatment effects and a case study of access to water in Bagezza sub-county.

6.4 Limitations

- Partial roll-out, selection bias, and matching

The primary limitation of this study is that end-line data collection after partial roll-out might have introduced selection bias. It it possible that, from the randomly assigned sub-counties, particular sub-counties were selected to be treated first and the treatment of other sub-counties was postponed. For example, the implementing partner may have started with sub-counties that are close to the capital for logistical reasons. Or the OPM may have treated politically favoured sub-counties first and other sub-counties later for political reasons. If sub-counties were selected for logistical or political reasons or due to other socio-economic characteristics, treatment is not random, our sample is not representative of the population we intended to analyse and some conclusions of this study may not be correct.

That is why we matched, ex ante, each treated sub-county to a control sub-county that is similar with regard to a range of observable pre-treatment characteristics. We used a range of sub-county characteristics that were likely to be known to OPM staff and may have affected how the intervention was rolled out. 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.

Classic matching attempts to reproduce the treatment group among the nontreated to re-establish experimental conditions in a nonexperimental setting and relies on observable variables to account for selection. We on the other hand deal with experimental conditions in an experimental setting, even though parts of our experiment were not implemented as planned. Our matching does not attempt to reproduce the treated among the nontreated but to select matching controls for our treated sub-counties. The aim of our and the classic matching method equal though: lineup comparison individuals according to sufficient observable characteristics to remove systematic differences in the evaluation outcome between treated and nontreated, so that the only remaining difference between the two groups is the treatment.

That is why both methods are subject to similar limitations. One main limitation of matching is related to data availability. We cannot be sure that the missing counterfactual, the matching control sub-county, exists in our sample. Some observations might not be matched perfectly, so that the estimated parameter is difficult to interpret (Blundell and Dias 2009). Another main limitation of matching is related to our ability to select the right information. We must observe and select the right characteristics to ensure that the unexplained share of the outcome is not related to the treatment decision. Heckman and Navarro (2004) show how on the one hand important and on the other hand difficult it is to select the appropriate set of variables for matching. If the conditioning set of variables is not right and complete, our estimates are biased (Blundell and Dias 2009). However, if observations are matched well and the right information is used, matching deals well with potential bias.

We acknowledge that the partial roll-out is a threat to internal validity that should not be ignored. At the same time, the fact that we started from a cluster randomized control design still provides a substantial advantage over studies that are based on observational data. For instance, potential selection emanating from partial roll-out is restricted to the sub-sample of sub-counties that were assigned to be treated group only, significantly reducing the scope for bias. In addition, the list of sub-counties to be treated that we shared with the implementing partner was organized by treatment group (information, deliberation and combined treatment), and in each

treatment group, sub-counties were listed alphabetically. Looking at the list in light of the partial roll-out, we get the impression that the OPM started at the top of the list and worked their way down. As a result, relatively more sub-counties that were assigned the information treatment were treated, and sub-counties toward the top of the list are also more likely to be treated. This pattern is confirmed when regressing the likelihood of being treated on the rank of the sub-county. This also suggests that OPM officials did not deliberately select certain sub-counties.

- Gender

Our impact evaluation was not designed to answer particular questions about gender, and this is a limitation of the study. While women and men were equally able to participate in barazas, we do not have data on who came to the events, and cannot say whether women and men were equally present. However, our facilitators were trained to encourage paying attention to the voice of women and minorities during meetings.

As women's priorities might have been raised and addressed less/more often in meetings, we compared female and male perceptions and prioritizations. Enumerators were instructed to interview household heads, so that 4714 (or 38 percent) were female respondents and 7831 (or 62 percent) were male respondents in our baseline data. Conditioning on the gender of the respondent, we do not find differences between women's and men's priorities for ten of fourteen statements. Using the unpaired two-samples t-test to compare the means of men and women, we found that men's perception is significantly different from women's perception for the statements:

a. Staff at government health centres or hospitals are rude to patients.
b. Children's learning outcomes are poor.
c. Agricultural inputs supplied by the government are of poor quality.
d. Agricultural extension agents are not aware of enterprises or agricultural inputs relevant for farmers.

For all four statement, women agreed significantly less than men, indicating that they are less concerned about the issues. Because we cannot find an issue that is prioritized more by women than by men (about which women are significantly more concerned than men), we cannot study whether such an issue was given more or less attention during and after the baraza. However, these differences and the lack thereof do not align with our expectations. It seems strange that women do not prioritize any issue more than men. There are several potential explanations for this surprising finding.

Firstly, prior studies in different contexts find that women respond in a more socially desirable fashion than men (Bernardi 2006, Chung and Monroe 2003, Hebert et al. 1995). Social desirability is the tendency of an individual to avoid criticism and to convey an image in line with social norms (Hebert et al. 1995). The social desirability response bias refers to the tendency of individuals to over-report socially desirable aspects and under-report undesirable aspects (Zerbe and Paulhus 1987). If respondents perceive it as socially desirable not to agree with statements that point to problems in public service delivery, i.e. not to criticize their government, this bias could explain why women responded to agree less than men.

Secondly, the result could be driven by selection bias: the women in our sample are often single household heads who are not necessarily representative of rest of the female population.

Thirdly, this result could indicate that the way we define women's priorities and perceptions is debatable. However, our study was not designed to answer particular gender related questions

and we therefore lack a better way to find out what women perceive to be important. That is why we cannot sufficiently test whether issues that are prioritized by women are more or less likely to be addressed during and after barazas.

7. Conclusions and recommendations

To improve governance and public service delivery, the Government of Uganda organizes community advocacy 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 these forums. 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) what is the impact of district level barazas. Baseline data on more than 12,500 households spread over almost 250 sub-counties in about 40 districts throughout Uganda was collected and OPM started implementing barazas following our 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 end-line data after partial roll-out. Various strategies were followed to diagnose, and reduce the consequences of, potential selection bias introduced by this partial roll out.

To answer the four questions mentioned above, we analysed a set of carefully selected variables, declared in a pre-registered analysis plan, and combined in indices. In this confirmatory analysis, we focus on five indices corresponding to the four main sectors – agriculture, infrastructure, health and education – and one overall index. We do not find a significant impact of the baraza programme on overall public service delivery. There are some indications that sub-county level barazas affected the agricultural sector, but the difference is only significant at the 10 percent level.

While we do not find that the baraza impacts public service delivery in general, we do find a variety of interesting effects when we look beyond the indices and analyse individual outcomes. In this second part of our analysis, which is more exploratory in nature, we find that in the agricultural sphere, sub-county level barazas significantly increase access to extension. We also see an increase in the likelihood that farmers received improved seeds from the government. This is consistent with the positive effect sub-county barazas seem to have on the likelihood that farmer associations or groups are formed in the village, and an increase in such institutions that is assisted by NAADS/OWC. We also find some improvements in public school infrastructure after a sub-county baraza, and a small reduction in waiting time at the water source.

We assess whether citizens interact more with politicians and service providers due to the meetings and find mixed results. The baraza intervention does not affect the likelihood that citizens participate in elections. However, we do find that the information component of a sub-county level baraza reduces in-kind contributions but increases cash contributions. Furthermore, the baraza interventions changed citizens' perceptions of a range of problems.

The lack of significant impact of barazas on the public service delivery indices surprises us, especially because qualitative research prior to end-line data collection suggested real effects of the intervention (Van Campenhout et al. 2018). That is why we investigate potential

explanations and run a series of robustness checks. We find a variety of interesting effects when considering heterogeneity in the treatment effects. Firstly, the slow roll-out of the baraza programme introduces sub-county heterogeneity in the time that passed between treatment and end-line data collection. Our analysis suggests that the effects of the interventions dissipate over time as enthusiasm fades, plans are abandoned and promises forgotten. Secondly, we asked sub-county officials whether they remember that a baraza took place, used this variable to check for heterogeneous treatment effects and found generally larger impacts of the intervention. Officials who recall the baraza might be intrinsically more motivated or alternatively, the information we received from the OPM might be inaccurate and some sub-counties that were not treated were indicated as being treated. Thirdly, we consider heterogeneity related to remoteness as barazas may have different effects on households that live further way from the sub-county headquarters. There are indications that the intervention is particularly effective for households in more remote locations. Fourthly, we reran the analysis with the subset of households that indicated that they are aware of the concept of baraza. The results suggest that especially households that are well-informed and interested in public service provision cash in on the baraza.

These mixed results are puzzling, especially because previous qualitative research suggests a real impact of the baraza programme (Van Campenhout et al. 2018). The nature of our interventions might be an explanation for this. Different sub-counties face different challenges, so that different issues are discussed during baraza events. As a result, subjects may receive different treatments and not standardized treatments comparable to the ones given in RCTs in biophysics. That is why a focus on average treatment effects may fail to find significant effects, as the impact is averaged over many sub-counties that in reality received “different” barazas. A case study in Bagezza sub-county confirms these issues related to non-standardized treatments.

Because barazas are designed to affect a broad range of public service outcomes, comparing the cost-effectiveness of the different types of barazas does not result in one straightforward conclusion. For some outcomes, sub-county level interventions seem to be more cost-effective than district-level interventions, for other outcomes, the opposite holds. Similarly, in some cases, information barazas are more cost-effective while in other cases deliberation barazas are. However, baraza interventions have an impact on many households and are inexpensive, so that the rate of return is substantial even if treatment effects are small.

Taking the impact of the baraza intervention on individual public service delivery outcomes, the heterogeneity in the treatment effects, and our concerns regarding non-standardized treatments into account, we recommend baraza meetings even though they do not have a measurable effect on our pre-registered indices. We do not conclude that sub-county level barazas are more effective than district level barazas or the other way around. Barazas at the sub-county level seem to have a larger effect on some outcomes while barazas at the district level seem to affect other variables. We do not find many significant impacts of district level barazas, but this might be due to insufficient statistical power. We would therefore recommend a mix of both approaches. Furthermore, both, the information component and the deliberation component of a sub-county level baraza seem to be important. We hence recommend the implementation of full barazas, especially because they are only slightly more expensive than information and deliberation barazas. Finally, since our heterogeneity analysis suggests that the effects of the interventions dissipate over time, baraza meetings should not be held only once but several times, for instance every two years.

The primary limitation of this study is that end-line data collection after partial roll-out might have introduced selection bias. We acknowledge that this is a threat to internal validity but believe that the fact that we started from a cluster randomized control design still provides a substantial advantage over studies that are based on observational data. We also provide an extensive investigation into the possibility that the results (or lack thereof) are driven by selection bias.

8. Appendixes

Note: The tables in the appendixes show the district level baraza vs. sub-county level baraza comparison and not the district level baraza vs. no baraza (control) comparison.

Descriptive statistics and balance tables

In table A.1, we test for balance between the treatment groups at baseline following the initial design of the experiment. 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 six household members, and about 30 percent of sampled households live in a house with a thatched grass roof. 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 cannot reject the null that households in these two groups are similar for all but one of the characteristics in table A.1. We do find that, at baseline, households assigned to a sub-county level baraza live farther from the nearest all weather road, and this difference is significant at the five percent significance level. 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 A.1, we report differences between households that were exposed to a sub-county deliberation baraza and households that were not. For this treatment, we cannot reject balance on any of the variables. In the last column, we report differences in outcomes between households that 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, judged by the cluster robust standard errors (CRSE), the differences are significant at a 10 percent level. However, it is well-known that when the clusters are few in number (say 30 or less) the cluster robust standard error is downward biased and tends to over-reject the null of no effect. We indeed find that the differences are not significant when randomization inference is used. Overall, out of 40 comparisons, we find that two differences are significant at the 5 percent level and one is 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.

In section 4.3 we provide additional balance tests to investigate whether the partial roll-out of the intervention introduced selection bias.

During both baseline and end-line, we collected some data at a more aggregate level. We visited sub-county headquarters and interviewed one politician and one civil servant there. For completeness, we also provide a balance table for this data. Results are in table A.2 in the Appendix. Despite the small sample size, also here the various subgroups seem to be balanced on a range of characteristics.

Table A.1 - Orthogonality tests

	mean	sub-county baraza	information	deliberation	jurisdictional tier
Household size	6.324	0.021	0.304*	-0.003	0.246
		(0.142)	(0.133)	(0.125)	(0.248)
Age of the household head (years)	46.501	0.736	0.464	0.725	1.427
	(14.615)	(0.681)	(0.594)	(0.714)	(0.802)
Head of household is woman (1=yes)	0.191	0.012	-0.014	0.004	-0.013
	(0.393)	(0.014)	(0.013)	(0.015)	(0.016)
Head finished primary education (1=yes)	0.213	-0.007	-0.02	-0.003	-0.026
	(0.410)	(0.017)	(0.020)	(0.020)	(0.027)
Thatched grass roof (1=yes)	0.298	-0.001	0.009	-0.032	0.011
	(0.457)	(0.026)	(0.025)	(0.023)	(0.036)
Traditional mud wall (1=yes)	0.424	0.021	-0.025	0.038	-0.034
	(0.494)	(0.043)	(0.040)	(0.039)	(0.104)
Distance to nearest all weather road (km)	0.906	0.167*	0.106	0.147	-0.192
	(0.915)	(0.106)	(0.095)	(0.092)	(0.138)
Access to extension (1=yes)	0.108	0.002	0.004	0.007	0.009
	(0.310)	(0.014)	(0.012)	(0.014)	(0.017)
Village Health Team in village (1=yes)	0.854	0.000	0.006	0.025	0.07
	(0.353)	(0.031)	(0.026)	(0.026)	(0.036)
Number of children in public schools	2.478	0.044	0.165+	0.038	0.139
	(2.074)	(0.095)	(0.091)	(0.089)	(0.155)
Number of observations	12,545	5,193	10,241	10,241	4,949

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports differences (and standard errors below) between baseline characteristics of households that received a district level combined information and deliberation baraza, and those that did not receive any baraza; **, * and + denotes significance at the 1, 5 and 10 percent levels.

Table A.2 - Balance table for sub-county level data

	mean	sub-county baraza	information	deliberation	jurisdictional tier
frequency of executive committee meetings	0.983	-0.03+	0.00	-0.03+	0.003
	(0.128)	(0.016)	(0.002)	(0.016)	(0.02)
proportion of health budget that has not been received	17.236	-6.836	-3.403	1.168	-0.726
	(26.086)	(4.475)	(4.807)	(4.796)	(4.913)
lengths of other all-weather roads	65.244	-14.968	-16.807	-13.466	-8.139
	(69.357)	(12.366)	(10.76)	(12.133)	(11.795)
proportion of households with electricity	17.154	4.259	2.132	1.862	-5.823
	(19.552)	(3.349)	(2.556)	(3.25)	(4.192)
number of male crop extension staff/agents	0.913	0.259	-0.08	0.031	-0.098
	(0.583)	(0.167)	(0.119)	(0.136)	(0.076)
proportion of households using improved seeds	41.293	-0.964	0.167	2.171	-5.102
	(26.748)	(3.534)	(3.889)	(3.758)	(3.314)
number of HC2s	3.428	0.039	0.458	-0.3	-0.846
	(3.34)	(0.757)	(0.777)	(0.63)	(0.604)
number of nurses/nursing assistants in-place in HC2s	6.015	0.322	1.262	0.924	3.003**
	-3.734	(0.7)	(0.983)	(0.74)	(0.839)
student enrolment in government secondary schools	733.866	29.312	178.374	106.592	-72.357
	(694.694)	(142.906)	(178.474)	(145.499)	(97.122)
Number of observations	262	102	168	168	102

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

Table A.3 - Impact of baraza on participation in elections

	mean	sub-county baraza	infor-mation	delibe-ration	juris-dictional tier
Hold any political/traditional positions?	0.303	0.018	-0.028	-0.017	-0.051*
	(0.46)	(0.019)	(0.021)	(0.03)	(0.023)
Voted in LC1 elections?	0.926	-0.014	0.000	0.011	-0.017
	(0.261)	(0.017)	(0.013)	(0.014)	(0.019)
Voted in LC3 elections?	0.884	0.025	0.016	0.027	-0.014
	(0.32)	(0.028)	(0.02)	(0.024)	(0.031)
Voted in LC5 elections?	0.898	-0.002	0.011	0.004	0.008
	(0.302)	(0.024)	(0.017)	(0.025)	(0.029)
Voted in the Presidential elections?	0.932	-0.008	0.003	0.003	0.018
	(0.252)	(0.018)	(0.012)	(0.016)	(0.018)
Voted in Parliamentary election?	0.922	-0.011	0.001	0.002	0.032
	(0.269)	(0.022)	(0.014)	(0.019)	(0.027)
Voted in Party leaders elections?	0.752	-0.01	-0.043	-0.01	-0.003
	(0.432)	(0.039)	(0.033)	(0.039)	(0.056)
Political Participation Index	0.000	0.006	-0.034	0.004	-0.043
	(0.646)	(0.053)	(0.041)	(0.049)	(0.066)
Number of observations	6,700	2,390	4,266	4,266	2,379

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

Sub-county level analysis

In addition to household surveys, we conducted surveys with government officials as respondents. Obviously, sample sizes are much smaller here, and so results should be interpreted with this caveat in mind. In each sub-county, we interviewed two officials: the highest-ranking politician (the LC3) and the highest ranking civil servant (sub-county chief). Sometimes, the deputy was interviewed. We have 261 observations in this dataset.

Agriculture

We start again with agriculture. We report results in table A.4.1. As in previous tables, the first column shows sample averages, with standard deviations in brackets below. In the second column, we report differences in outcomes between sub-counties that received a typical sub-county level baraza and sub-counties that did not receive any baraza. In the third column, we report differences between outcomes of sub-counties where an information baraza was organized and outcomes of sub-counties that were not exposed to an information baraza. In the fourth column, we report differences between outcomes of sub-counties where a deliberation baraza was organized and outcomes of sub-counties that were not exposed to a deliberation baraza. Finally, in the fifth column, we directly compare sub-counties that received a sub-county level baraza to sub-counties that were exposed to district level baraza.

Government officials report that on average 14.3 percent of the agricultural budget was not received. We do not find evidence that the baraza intervention affected this percentage.

We then look at perceptions of problems in the agricultural sector by officials. Over the past year, officials received on average 2.9 complaints related to agricultural service provision. The number of complaints seemed to reduce after a sub-county level baraza took place. As in the household questionnaire, officials were also asked to rate their agreement with various statements. We do not find that barazas affect perceptions on input quality. However, officials in sub-counties with deliberation barazas report that there is increased transparency in how farmers are selected to receive agricultural inputs. Officials in sub-counties with information barazas agree less with the assertion that extension agents rarely visit. Officials in sub-counties with deliberation barazas are also more of the opinion that extension agents are aware of what their customers want. The above seems to suggest perceptions became more positive in the agricultural sector after sub-county level barazas, but it is unclear if it is the information or deliberation component that is driving this result.

Turning to outcomes, we investigate the effect of barazas on access to extension at home as reported by officials. Recall that when analysing the household data we found that significantly more households in areas that received a sub-county level baraza were visited by an expert at home. Analysing the responses of government officials, we find that the number of male crop extension agents is about one person higher in areas where a deliberation baraza took place. There is also a significant difference in the number of male crop extension agents when directly comparing sub-county level barazas to district level baraza, with more staff available after a district level baraza. The number of female crop extension staff/agents was not affected by the baraza intervention. We also find substantial reductions in the number of demonstration sites as a result of sub-county level barazas. This is surprising, given that in the household level data there is some evidence of increased visits to extension offices, demonstration sites and model farmers, especially after matching to reduce potential bias introduced by the partial roll-out (table A.5).

Looking that the use of modern inputs, we find that for both fertilizers and improved planting material, there is a negative and significant difference between areas exposed to a district level baraza and areas exposed to a sub-county level baraza. We also see that the percentage of households in the sub-county that reportedly used improved seed or fertilizer is higher in areas that received a sub-county baraza as opposed to control sub-counties, but the difference is not significant.

Using household survey data, we find that the proportion of households that received improved seed from the government extension system is significantly higher in areas where a sub-county level baraza took place. Asking government officials about the frequency of improved seed distribution, we do not find a significant difference between sub-counties with a sub-county level baraza and control sub-counties. However, we do see that the frequency of improved seed distribution is about 0.4 higher in areas where an information baraza took place. We also see that the frequency of improved breeds of cattle, goat, pig, poultry distribution is higher in areas that were exposed to a sub-county level baraza. Grievances related to the distribution of seed and livestock (goats and milk cows) were often encountered during qualitative work.

Table A.4.1 - Impact on agriculture (sub-county level analysis)

	mean	sub-county baraza	information	deliberation	juris-dictional tier
Political effort					
agricultural budget that has not been received‡ (in %)	14.284	-3.446	-1.991	7.389	-6.375
	(29.046)	(6.443)	(5.41)	(9.309)	(3.906)
Perception					
number of complaints‡	2.945	-2.874+	-1.565	-1.44	0.824
	(7.686)	(1.636)	(1.751)	(1.781)	(0.81)
"Agricultural inputs supplied by the government are of poor quality."	5.669	0.254	-0.24	1.088	-0.313
	(3.179)	(0.666)	(0.659)	(0.867)	(0.667)
"Lack of transparency in how farmers are selected to receive ag inputs from govt."	5.225	-0.76	0.054	-1.423+	0.201
	(3.244)	(0.611)	(0.684)	(0.781)	(0.886)
"Agricultural extension agents rarely visit."	5.199	0.895	-1.355+	-1.063	-0.313
	(3.194)	(0.665)	(0.76)	(0.766)	(0.469)
"Agricultural extension agents are not aware of needs relevant to farmers."	4.483	-0.125	-0.792	-1.527+	-0.402
	(3.069)	(0.557)	(0.72)	(0.799)	(0.41)
Outcomes					
number of male crop extension agents‡	1.123	-0.207	0.248	1.047+	0.205*
	(1.233)	(0.133)	(0.309)	(0.517)	(0.089)
number of female crop extension agents‡	0.36	0.382	0.206	0.357	-0.321
	(1.049)	(0.236)	(0.235)	(0.39)	(0.24)
number of demonstration sites‡	3.157	-2.249*	-1.276	-2.461*	2.977*
	(6.235)	(1.029)	(1.257)	(1.071)	(1.244)
HH using purchased fertilizers‡ (in %)	28.11	5.129	-1.79	-8.809	-19.808*
	(29.98)	(7.588)	(7.367)	(7.905)	(7.195)
HH using improved seeds‡ (in %)	44.136	5.288	5.393	-2.399	-14.907*
	(29.9)	(5.316)	(6.45)	(7.193)	(5.63)
HH using pesticides/herbicides/fungicides‡ (in %)	46.195	3.096	7.286	-9.217	-5.65
	(29.289)	(6.205)	(5.916)	(7.779)	(7.302)
HH using improved livestock breeds‡ (in %)	23.131	2.499	0.394	-7.672	-15.533
	(25.196)	(10.283)	(6.568)	(6.279)	(11.049)
frequency of improved seed distribution‡	1.932	0.000	0.393+	0.496	0.879*
	(1.776)	(0.223)	(0.219)	(0.293)	(0.378)
frequency of improved breed distribution	1.053	0.602+	0.097	0.492	-0.668*
	(1.112)	(0.33)	(0.179)	(0.328)	(0.282)
frequency of fertilizer/manure distribution	0.524	-0.1	0.284	-0.215	0.316
	(1.399)	(0.178)	(0.542)	(0.201)	(0.291)
frequency of pesticide/herbicide/fungicide distribution	0.557	0.634	-0.036	-0.174	-0.447
	(1.206)	(0.389)	(0.156)	(0.185)	(0.408)
Number of observations	262	102	168	168	102

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports the effect (and standard errors below) of the administrative placement of the baraza intervention. **, * and + denotes significance at the 1, 5 and 10 percent levels. ‡ indicates that missing observations were interpreted as zeros. † indicates that we did not control for the baseline value.

Infrastructure

We now turn to infrastructural related outcomes as reported by sub-county officials (table A.4.2). We start with perceptions. Also for infrastructure, we recorded the number of complaints. We find a reduction of -5.1 water related complaints after a sub-county level baraza. It seems that the deliberative component is the main driver behind this result. Moreover, government officials were asked to report their agreement with two water infrastructure related statements. When asked whether “Access to a drinking water source is a serious problem”, government officials in sub-counties that received a sub-county level baraza agreed significantly more with this statement. We find a similar effect for the statement: “Drinking water is usually dirty”. Perhaps, sub-county level baraza made officials more sensitive to this issue.

We also include some questions related to road infrastructure. Looking at the household data, we do not find that the baraza programme reduces the average distance of households to the nearest all weather road. This is in line with our findings from surveying government officials. However, when officials were asked to report their agreement with the statement “Availability/ Access to all-weather roads is a serious problem.”, officials in sub-counties that received a typical sub-county level baraza agreed significantly less with this statement.

In the household level analysis, we learned that the difference in distance to the primary water source during the dry season is never significantly different from zero, but that there is a significant reduction in the time that one has to wait at the water source in areas that were exposed to the sub-county level baraza intervention. Government officials reported on a range of different water sources. We find that there are 14 more boreholes in sub-counties where an information baraza was organized. We further find a reduction of the number of protected springs in sub-counties that were exposed to a sub-county level baraza. At the same time, we see an increase in the number of protected springs in areas that were exposed to a district level baraza, resulting in a difference of 7.3 which is significant at the 1 percent level.

The increase in the number of boreholes in sub-counties where an information baraza was organized and the decrease in the number of protected springs in sub-counties that received a sub-county level baraza could be the reason why the difference in distance to the primary water source of households during the dry season is never significantly different from zero because the two effects offset each other. Because the positive impact on the number of boreholes is much larger than the negative impact on the number of protected springs, this could be an explanation for the significant reduction in time that households report to wait at the water source in areas that were exposed to the sub-county level baraza intervention.

Table A.4.2 - Impact on infrastructure (sub-county level analysis)

	mean	sub-county baraza	information	deliberation	jurisdictional tier
Perception					
number of complaints‡	4.966	-5.144*	-0.943	-5.17+	1.12
	(9.145)	(2.449)	(2.256)	(2.51)	(1.267)
"Access to a drinking water source is a serious problem."	5.953	1.623*	0.203	0.702	-0.854
	(3.057)	(0.719)	(0.688)	(0.817)	(0.697)
"Drinking water is usually dirty."	5.025	1.729*	-0.164	1.166	-2.373*
	(3.116)	(0.687)	(0.675)	(0.914)	(0.85)
"Availability/Access to all-weather roads is a serious problem."	6.784	-1.137+	-0.502	0.177	-0.296
	(2.774)	(0.558)	(0.542)	(0.648)	(0.887)
Outcomes					
lengths of tarmac roads‡	3.393	-4.317	-4.182	-5.61+	-0.988
	(9.545)	(2.817)	(2.578)	(3.257)	(0.678)
lengths of other all-weather roads‡	50.255	8.485	1.654	40.143	6.108
	(74.271)	(11.343)	(10.359)	(30.295)	(12.236)
number of boreholes‡	16.763	2.601	14.467**	7.371	1.133
	(26.371)	(2.843)	(5.079)	(4.808)	(3.602)
number of protected springs‡	9.275	-5.287+	1.943	1.918	7.334**
	(19.549)	(2.767)	(3.304)	(4.727)	(2.1)
number of protected dug/shallow wells‡	3.585	-2.046	-0.152	1.824	2.462+
	(7.965)	(1.729)	(1.397)	(3.116)	(1.268)
number of unprotected dug/shallow wells‡	8.513	-0.094	1.885	8.02	-4.073
	(18.575)	(3.589)	(3.569)	(7.406)	(3.315)
number of piped/gravity flows†	12.234	7.473	2.04	-4.963	-18.955+
	(37.886)	(11.924)	(13.147)	(7.795)	(10.346)
number of observations	262	102	168	168	102

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports the effect (and standard errors below) of the administrative placement of the baraza intervention. **, * and + denotes significance at the 1, 5 and 10 percent levels. ‡ indicates that missing observations were interpreted as zeros. † indicates that we did not control for the baseline value.

Health

We now study health related outcomes and report them in table A.4.3. We find that, on average, 14.1 percent of the health budget has not been received and that this proportion is independent of treatment groups. While we see that the number of health related complaints reduces for most comparisons, the only significant difference is the one between sub-counties with a deliberation baraza and sub-counties without, the former receiving on average 2 complaints less. For perceptions, absenteeism seems to be less of a problem in areas that experience a district level baraza than in areas that were exposed to a sub-county level baraza.

We then investigate outcomes related to VHTs. Using the household data we find that the share of households that reports that a VHT is present in their village and the likelihood that individuals participate as VHT members is not impacted by the baraza intervention. Responses of government officials are line with household responses. In fact, the number of VHT members

reduces by 14.6 in sub-counties with a sub-county level baraza. The number of VHT members is independent of the other treatment groups.

Using household data, we found that access to public health facilities was independent of the treatment groups. Also, we did not find that barazas reduced the distance to the nearest government health facility. Here, we look at the number of health centres in the sub-counties as reported by the officials. The number of Health Centres 2 (HC2s) is not significantly different for comparisons of the different sub-county level barazas. However, when directly comparing sub-counties that received a sub-county level baraza to sub-counties that were exposed to district level baraza, we do find a significant increase of 0.3 centers. Furthermore, the number of Health Centers 3 (HC3s) increases by 0.3 in sub-counties that were exposed to a sub-county level baraza. Both are significant at the 5 percent level.

After looking at the number of HC2s and HC3s, we want to take a closer look at the situation inside these public health facilities. We start with strafing in HC2s, differentiating between clinical officers, nurses and birth attendants. Interestingly, we find that nurses and birth attendants are present in higher numbers after a district level barazas then after a sub-county level barazas. In fact, there are some indications that a sub-county level baraza leads to a reduction in staff.

Equipment of HC2s matters for service delivery. The number of HC2s with a safe drinking water source, with laboratory tests, with a medical waste pit, or with Human Immunodeficiency Viruses (HIV)/Acquired Immune Deficiency Syndrome (AIDS) guidance and counselling services is independent of the treatment groups. In sub-counties that received a sub-county level baraza, the number of HC2s with immunization facilities, the number of HC2s with out-patient services, and the number of HC2s with family planning services all increase, while the number of HC2s with in-patient care reduces. Administrative placement also seems to matter for equipment of HC2s. If we compare sub-counties that received a sub-county level baraza to sub-counties that were exposed to district level baraza, there are 0.1 more HC2s with electricity, 0.4 percentage points more with staff houses for all relevant employees and 0.2 more with in-patient care in areas that were exposed to a higher level baraza.

We also look at staffing in HC3s. The number of doctors, clinical officers, medical assistants, nurses, nursing assistants and laboratory technicians on payroll in HC3s are all independent of the treatment groups. The number of midwives on payroll in HC3 is lower in sub-counties that received a district level baraza compared to sub-counties that were exposed to sub-county level baraza. The numbers of in-patient care beds in HC3s, the number of HC3s with electricity and with a mortuary/cold room are independent of the treatment groups. In sub-counties that received a sub-county level baraza, HC3s are more likely to have a safe drinking water source, are more likely to provided laboratory tests, are more likely to provide immunization services, provide more out-patient services, provide more family planning services and antenatal care. Sub-counties that received a baraza are also more likely to have appropriate medical waste disposal facilities and offer HIV/AIDS guidance and counselling services. These outcomes seem to be mostly driven by the information component.

Table A.4.3 - Impact on health sector (sub-county level analysis)

	mean	sub-county baraza	infor-mation	delibe-ration	juris-dictional tier
Political effort					
proportion of health budget that has not been received‡	14.089	-1.429	8.521	5.522	-6.923
	(28.051)	(6.472)	(6.868)	(9.498)	(4.753)
Perception					
number of complaints‡	1.75	-1.641	-1.016	-1.986+	0.545
	(5.719)	(1.363)	(1.285)	(1.115)	(0.623)
"Access to a government health centre or hospital is a serious problem."	6.64	-0.426	0.194	-0.761	-0.378
	(3.026)	(0.895)	(0.62)	(0.657)	(0.781)
"Government health centres or hospitals do not have relevant medicines."	6.826	0.553	-0.127	0.505	-0.434
	(2.818)	(0.52)	(0.644)	(0.627)	(0.37)
"Staff at government health centres or hospitals are rude to patients."	4.394	0.797	-0.047	-0.664	-0.597
	(2.821)	(0.605)	(0.61)	(0.66)	(0.739)
"Medical staff at government health centres or hospitals are often absent."	4.411	0.757	-0.049	-0.976	-1.683*
	(2.943)	(0.565)	(0.587)	(0.685)	(0.666)
Outcomes					
number of villages with VHTs‡	43.225	-8.924	-4.327	-1.829	-1.831
	(24.644)	(6.351)	(6.282)	(6.739)	(3.768)
number of VHT members‡	73.907	-14.644+	-5.301	11.937	-2.738
	(40.67)	(7.936)	(8.061)	(11.889)	(10.935)
number of HC2s‡	1.097	-0.124	0.112	0.14	0.34+
	(1.229)	(0.181)	(0.154)	(0.162)	(0.153)
number of clinical officers on payroll in HC2s‡	0.343	-0.174	-0.052	0.024	0.135
	(0.925)	(0.22)	(0.148)	(0.262)	(0.121)
number of nurses/nursing assistants on payroll in HC2s‡	2.047	-1.273*	0.002	-0.873	1.409**
	(3.103)	(0.49)	(0.493)	(0.594)	(0.417)
number of birth attendants on payroll in HC2s‡	0.445	-0.32*	0.088	0.615	0.269**
	(1.142)	(0.146)	(0.165)	(0.478)	(0.072)
number of HC3s‡	0.801	0.307*	0.333*	-0.016	-0.292
	(0.67)	(0.125)	(0.127)	(0.117)	(0.208)
number of in-patient care beds in HC3s‡	5.839	-0.451	0.472	0.01	-1.609
	(7.359)	(1.6)	(1.721)	(1.871)	(1.636)
number of doctors on payroll in HC3s†	0.063	-0.048	-0.015	-0.044	0.053
	(0.312)	(0.096)	(0.097)	(0.118)	(0.066)
number of clinical officers on payroll in HC3s‡	1.042	0.193	0.228	0.164	-0.238
	(1.122)	(0.224)	(0.271)	(0.297)	(0.257)
number of medical assistants on payroll in HC3s‡	0.39	-0.158	-0.048	-0.104	0.012
	(0.937)	(0.18)	(0.172)	(0.226)	(0.202)
number of nurses/nursing assistants on payroll in HC3s‡	2.826	0.579	1.063	0.951	-0.625
	(3.072)	(0.687)	(0.826)	(0.997)	(0.793)
number of midwives on payroll in HC3s‡	1.691	-0.194	-0.063	-0.529	-0.745**
	(2.205)	(0.403)	(0.432)	(0.458)	(0.218)
number of laboratory technicians on payroll in HC3s‡	0.936	0.497	0.339	-0.236	-0.508

	(1.126)	(0.309)	(0.251)	(0.227)	(0.293)
number of HC2s with electricity‡	0.237	0.243	-0.183+	0.151	0.136*
	(0.533)	(0.24)	(0.104)	(0.156)	(0.056)
number of HC2s with safe drinking water source‡	0.564	0.738	0.122	0.246	-0.394
	(1.084)	(0.47)	(0.21)	(0.251)	(0.266)
number of HC2s with staff houses for all relevant employees‡	0.72	0.945	0.135	0.17	0.359+
	(1.587)	(0.642)	(0.229)	(0.145)	(0.16)
number of HC2s with laboratory tests‡	0.53	-0.128	-0.007	-0.313	0.268
	(1.008)	(0.086)	(0.178)	(0.197)	(0.204)
number of HC2s with immunization facilities‡	0.886	0.823**	0.117	-0.139	0.26
	(1.156)	(0.264)	(0.166)	(0.174)	(0.263)
number of HC2s with in-patient care‡	0.242	-0.169+	-0.071	-0.089	0.235*
	(0.712)	(0.088)	(0.11)	(0.107)	(0.103)
number of HC2s with out-patient services‡	1	0.661*	0.221	0.187	0.034
	(1.203)	(0.243)	(0.134)	(0.172)	(0.161)
number of HC2s with family planning services‡	0.911	0.576*	0.098	0.136	0.113
	(1.169)	(0.238)	(0.155)	(0.207)	(0.188)
number of HC2s with medical waste pit‡	0.826	0.401	0.33	0.08	0.326
	(1.174	(0.355)	(0.205)	(0.233)	(0.362)
number of HC2s with HIV/AIDS guidance and counselling‡	0.856	0.375	-0.018	-0.305	0.092
	(1.169)	(0.32)	(0.176)	(0.216)	(0.226)
number of HC3s with electricity‡	0.525	0.299	0.13	0.009	-0.04
	(0.635)	(0.402)	(0.139)	(0.163)	(0.098)
number of HC3s with safe drinking water source‡	0.568	0.997*	0.216	0.005	-0.116
	(0.632)	(0.176)	(0.141)	(0.146)	(0.136)
number of HC3s with staff houses for all relevant employees‡	0.525	-0.025	-0.203	-0.226	-0.289*
	(1.128)	(0.752)	(0.266)	(0.228)	(0.103)
number of HC3s with laboratory tests‡	0.797	0.827**	0.165	-0.133	0.113
	(0.821)	(0.106)	(0.155)	(0.139)	(0.257)
number of HC3s with immunization facilities‡	0.818	0.659**	0.463*	0.022	-0.059
	(0.838)	(0.127)	(0.199)	(0.114)	(0.299)
number of HC3s with in-patient care‡	0.674	0.037	0.29+	-0.066	0.177
	(0.69)	(0.14)	(0.158)	(0.169)	(0.206)
number of HC3s with out-patient services‡	0.775	0.632*	0.317**	0.022	-0.089
	(0.694)	(0.136)	(0.105)	(0.12)	(0.314)
number of HC3s with family planning services‡	0.839	0.786*	0.125	-0.141	-0.066
	(0.825)	(0.097)	(0.162)	(0.149)	(0.31)
number of HC3s with antenatal care‡	0.792	0.659**	0.275*	0.000	-0.032
	(0.687)	(0.127)	(0.105)	(0.119)	(0.289)
number of HC3s with maternity wards‡	0.763	0.221	0.343+	-0.054	0.047
	(0.705)	(0.179)	(0.171)	(0.171)	(0.272)
number of HC3s with placenta pit‡	0.725	0.608*	0.158	-0.038	0.104
	(0.669)	(0.135)	(0.104)	(0.134)	(0.248)
number of HC3s with medical waste pit‡	0.742	0.66*	0.254*	0.035	0.019

	(0.712)	(0.131)	(0.113)	(0.144)	(0.192)
number of HC3s with HIV/AIDS guidance and counselling‡	0.78	0.586*	0.322**	0.073	-0.058
	(0.71)	(0.145)	(0.109)	(0.122)	(0.26)
number of HC3s with mortuary/cold room‡	0.097	0.018	0.217	0.005	-0.011
	(0.572)	(0.065)	(0.173)	(0.061)	(0.072)
number of HC3s with isolation room for special cases‡	0.191	0.136	0.29+	-0.023	0.000
	(0.805)	(0.143)	(0.171)	(0.072)	(0.176)
number of observations	262	102	168	168	102

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports the effect (and standard errors below) of the administrative placement of the baraza intervention. **, * and + denotes significance at the 1, 5 and 10 percent levels. ‡ indicates that missing observations were interpreted as zeros. † indicates that we did not control for the baseline value.

Education

Now we assess the impact of the baraza intervention on education outcomes as reported by sub-county government officials. Like other sectors, about 14 percent of the budget has not been received and this proportion did not change as a result of the baraza intervention. The number of education related complaints also remains stable over subgroups.

We find that student enrolment in government primary schools and student enrolment in government secondary schools are also not affected by the baraza intervention, which is in line with the household level analysis. However, there seems to be an effect on dropout rates. The dropout rate for girls in primary schools is 13.2 percentage points lower in sub-counties where an information baraza was organized and 12.1 percentage points lower in sub-counties that received a district level baraza, compared to sub-counties that were exposed to sub-county level baraza. Also, the dropout rate for boys in primary schools is 6.8 percentage points lower in sub-counties with an information baraza, compared to sub-counties without.

From the household data, we learned that the baraza program did not have an impact on the distance to a government operated primary or secondary school. Looking at the government officials' data, we see that the baraza program did not affect the number of government primary schools or the number of government secondary schools, which could explain why the baraza program did not impact the distance to government operated schools.

The number of teachers on payroll in government primary schools in not affected by the baraza intervention. The number of teachers on payroll in government secondary schools is: There are on average 8.6 more secondary school teachers on payroll in government secondary schools in sub-counties that received a sub-county level baraza and this effect seems to come mostly from the participation component. Comparing sub-counties that received a sub-county level baraza to sub-counties that were exposed to district level baraza, we find that secondary schools in the latter group had significantly less teachers. However, care needs to be taken when interpreting these results because of a limited number of observations.

According to the household level data, the baraza intervention does not seem to affect whether the school has a SMC. In line with this result, the number of government schools (primary or secondary) with an active SMC is not affected by the baraza intervention, according to government officials.

Finally, government officials were asked about their opinion on four problems that were often mentioned by stakeholders. The intervention does not significantly affect agreement with the statements “Access to a government primary school is a serious problem” and “Children’s learning outcomes in government schools are poor”. When government officials were asked whether “Teachers in government schools are often absent”, they agreed significantly more in sub-counties that received a sub-county level baraza. Officials in sub-counties that were exposed to district level barazas were less of the opinion that absenteeism was a problem than officials in sub-counties that were exposed to district level barazas.

Table A.4.4 - Impact on education sector (sub-county level analysis)

	mean	sub-county baraza	infor-mation	delibe-ration	juris-dictional tier
Political effort					
number of government schools with active SMC‡	7.398	-0.986	1.595	-3.395	-1.913
	(7.595)	(2.005)	(1.749)	(2.029)	(1.529)
proportion of education budget that has not been received‡	13.881	1.339	6.384	4.831	-2.378
	(27.959)	(6.539)	(5.463)	(7.896)	(3.619)
Perception					
number of complaints‡	2.411	-0.247	1.190	-0.24	-0.441
	(4.495)	(0.929)	(1.139)	(0.883)	(0.644)
"Access to a government primary school is a serious problem."	4.225	-0.380	0.203	-1.054	-0.496
	(2.996)	(0.708)	(0.683)	(0.895)	(0.441)
"Teachers in government schools are often absent."	4.318	1.119+	0.331	0.63	-1.96**
	(2.767)	(0.654)	(0.536)	(0.639)	(0.556)
"Children's learning outcomes in government schools are poor."	7.542	-0.256	0.403	-0.005	-0.077
	(2.383)	(0.533)	(0.556)	(0.612)	(0.553)
Outcomes					
dropout rate for girls in primary schools	35.045	1.085	-13.228**	-8.808	-12.104+
	(22.047)	(6.169)	(4.478)	(5.346)	(5.658)
dropout rate for boys in primary schools	26.247	-0.088	-6.827+	-6.069	-6.938
	(19.533)	(4.799)	(4.013)	(4.849)	(4.564)
number of government primary schools‡	8.737	-1.099	-0.443	0.194	-0.747
	(6.31)	(1.008)	(1.039)	(1.09)	(1.095)
student enrolment in government primary schools‡	4458.78	-2860.48	-531.81	-2814.12	341.71
	(9512.47)	(2952.79)	(2561.43)	(3333.18)	(572.89)
number of teachers on payroll in government primary schools‡	58.386	6.690	17.379	9.720	5.919
	(60.045	(11.897)	(15.201)	(16.079)	(9.742)
number of government secondary schools‡	0.996	-0.328	0.163	-0.320	0.064
	(1.472)	(0.314)	(0.342)	(0.339)	(0.121)
student enrolment in government secondary schools‡	388.453	81.974	67.362	27.858	-41.100
	(546.069)	(147.351)	(116.486)	(125.191)	(122.726)
number of teachers on payroll in government secondary schools‡	8.737	8.616*	0.313	6.141+	-8.896*
	(13.826)	(3.963)	(2.633)	(3.487)	(2.969)
Number of observations	262	102	168	168	102

Note: First column reports sample means (and standard deviations below); Column 2 reports effect (and standard errors below) of the sub-county level baraza intervention; Column 3 reports the effect (and standard errors below) of the information component of the baraza intervention; Column 4 reports the effect (and standard errors below) of the deliberation component of the baraza intervention; Column 5 reports the effect (and standard errors below) of the administrative placement of the baraza intervention. **, * and + denotes significance at the 1, 5 and 10 percent levels. ‡ indicates that missing observations were interpreted as zeros. † indicates that we did not control for the baseline value.

Table A.5 - Impact of baraza on agricultural outcomes (matched Analysis of variance (ANOVA))

	mean	sub-county baraza	infor-mation	delibe-ration	juris-dictional tier
Household used inorganic fertilizers?†	0.314	0.004	0.061	-0.03	-0.021
	(0.464)	(0.038)	(0.053)	(0.066)	(0.044)
Household used improved seed? †	0.441	0.006	-0.019	-0.097+	-0.106+
	(0.497)	(0.043)	(0.058)	(0.048)	(0.052)
Received improved seeds from govt?	0.146	-0.001	0.011	0.024	-0.063+
	(0.353)	(0.032)	(0.041)	(0.056)	(0.032)
Household used agro-chemicals?	0.577	0.048	0.011	-0.007	-0.042
	(0.494)	(0.057)	(0.058)	(0.06)	(0.06)
Household used improved livestock inputs?	0.27	0.07	0.053	-0.006	-0.046
	(0.444)	(0.044)	(0.046)	(0.043)	(0.042)
Did an agricultural expert visit your home? †	0.212	0.054+	0.023	0.038	-0.107**
	(0.409)	(0.031)	(0.05)	(0.064)	(0.029)
Visited extension office/demo site/model farm? †	0.306	0.077+	0.087	0.036	-0.088*
	(0.461)	(0.043)	(0.053)	(0.062)	(0.038)
Are officials aware of extension demand?	0.832	0.024	0.03	-0.021	-0.006
	(0.374)	(0.03)	(0.032)	(0.034)	(0.033)
Not consulted for extension content?	0.295	0.007	0.032	-0.005	-0.083+
	(0.456)	(0.032)	(0.042)	(0.048)	(0.04)
Are farmer associations/groups in this village?	0.394	0.012	-0.04	-0.033	-0.108+
	(0.489)	(0.052)	(0.036)	(0.042)	(0.056)
Farmer groups supported by govt? †	0.381	0.090*	0.07	0.073	-0.085
	(0.486)	(0.04)	(0.059)	(0.053)	(0.048)
Received help in marketing from govt? †	0.194	0.082+	0.037	0.013	-0.099*
	(0.396)	(0.04)	(0.049)	(0.051)	(0.04)
Received help in marketing from coop? †	0.073	0.029	-0.006	0.003	0.011
	(0.26)	(0.028)	(0.023)	(0.023)	(0.029)
Number of observations	6,703	666	1,568	1,584	1,517

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

Table A.6 - Impact of baraza on infrasctructure (matched ANOVA)

	mean	sub-county baraza	infor-mation	delibe-ration	juris-dictional tier
Household uses unprotected water source†	0.195	-0.006	0.035	0.036	-0.056
	(0.396)	(0.051)	(0.055)	(0.047)	(0.057)
Distance to water source (km) †	0.778	-0.056	-0.036	-0.061	0.031
	(0.572)	(0.06)	(0.058)	(0.076)	(0.069)
Waiting time at source (min) †	3.188	0.002	-0.003	-0.303	0.126
	(1.638)	(0.207)	(0.18)	(0.227)	(0.145)
Is there a Water User Committee in the village?†	0.579	0.043	0.056	0.007	-0.005
	(0.494)	(0.06)	(0.046)	(0.057)	(0.061)
Is member of Water User Committee?	0.168	0.080**	0.018	0.063*	-0.047
	(0.374)	(0.026)	(0.029)	(0.025)	(0.035)
Water User Committee holds public meetings?	0.431	0.028	0.051	0.044	0.012
	(0.495)	(0.057)	(0.047)	(0.056)	(0.051)
Satisfied with quality of drinking water?	0.594	0.02	0.043	-0.095	-0.006
	(0.491)	(0.066)	(0.058)	(0.058)	(0.05)
Treat water before drinking? (boil or treat)	0.593	-0.005	-0.08	-0.024	-0.028
	(0.491)	(0.054)	(0.059)	(0.055)	(0.071)
Distance to nearest all weather road (km) †	3.102	0.211	-0.167	-0.268	-0.262
Number of observations	6,703	578	1,461	1,440	1,400

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

Table A.7 - Impact of baraza on the health sector (matched ANOVA)

	mean	sub-county baraza	infor-mation	delibe-ration	juris-dictional tier
Seek treatment for fever in public health facility †	0.696	-0.018	0.015	0.007	0.061
	(0.460)	(0.046)	(0.044)	(0.051)	(0.076)
Go to public health facility to give birth†	0.828	-0.007	-0.013	-0.016	-0.046
	(0.377)	(0.048)	(0.039)	(0.043)	(0.069)
Is there a VHT in village? †	0.891	0.007	0.034	0.033	-0.026
	(0.312)	(0.033)	(0.024)	(0.034)	(0.027)
Member of VHT?	0.127	0.007	0.001	0.002	-0.052+
	(0.333)	(0.026)	(0.024)	(0.019)	(0.024)
VHT organizes any public meetings?	0.415	0.090+	0.031	0.025	-0.103+
	(0.493)	(0.051)	(0.047)	(0.053)	(0.052)
Distance to nearest govt health facility (km) †	4.033	0.149	-0.144	-0.172	-0.341
	(1.283)	(0.202)	(0.263)	(0.241)	(0.318)
Any members sick?	0.646	0.025	0.004	0.061	0.018
	(0.478)	(0.036)	(0.042)	(0.048)	(0.036)
Number of days ill?	2.486	0.044	-0.148	-0.04	0.033
	(2.157)	(0.190)	(0.214)	(0.260)	(0.125)
Number of days school/work missed due to illness†	2.176	0.08	-0.086	0.023	0.031
	(1.987)	(0.166)	(0.180)	(0.239)	(0.098)
Waiting time before being attended (min) †	4.763	-0.128	-0.24	-0.278+	0.014
	(0.987)	(0.107)	(0.152)	(0.144)	(0.112)
Has visited traditional health practitioner? †	0.283	-0.051	0.013	0.044	-0.001
	(0.450)	(0.049)	(0.039)	(0.036)	(0.037)
Patient was examined by in-charge/doctor	0.432	0.093	0.087	-0.088	-0.056
	(0.496)	(0.061)	(0.052)	(0.066)	(0.054)
Time of examination	3.415	0.054	-0.022	0.001	-0.12
	(0.758)	(0.095)	(0.122)	(0.099)	(0.092)
Paid anything	0.2	0.033	0.004	0.025	-0.018
	(0.401)	(0.039)	(0.036)	(0.061)	(0.032)
Received meds in hospital	0.677	0.021	-0.038	0.053	0.024
	(0.468)	(0.041)	(0.037)	(0.040)	(0.064)
Had to buy meds outside of hospital	0.955	0.002	-0.007	0.015	-0.042*
	(0.207)	(0.026)	(0.046)	(0.035)	(0.017)
Satisfied with services at hospital	0.642	0.047	-0.078	-0.069	-0.055
	(0.480)	(0.039)	(0.048)	(0.053)	(0.036)
Number of observations	6,703	326	786	789	771

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

Table A.8 - Impact of baraza on education (matched ANOVA)

	mean	sub-county baraza	information	deliberation	jurisdictional tier
Number of children in UPE or USE†	1.696	0.253	-0.078	0.045	0.041
	(1.842)	(0.159)	(0.143)	(0.136)	(0.143)
Distance to public school (km) †	1.424	-0.018	-0.031	-0.055	0.066
	(0.707)	(0.114)	(0.118)	(0.09)	(0.081)
Has complete boundary fence? †	0.416	0.106	-0.085	-0.096	-0.054
	(0.493)	(0.07)	(0.075)	(0.062)	(0.087)
Has electricity?	0.352	0.195**	0.004	-0.045	-0.091
	(0.478)	(0.058)	(0.052)	(0.061)	(0.054)
Has water facility? †	0.677	0.094	-0.029	0.032	-0.103*
	(0.468)	(0.067)	(0.079)	(0.082)	(0.039)
Has PTA?	0.959	0.005	-0.033	-0.054	0.003
	(0.198)	(0.01)	(0.023)	(0.049)	(0.016)
Has School Management Committee? †	0.934	0.050+	-0.011	-0.054	0.027
	(0.248)	(0.029)	(0.027)	(0.052)	(0.037)
Informed about SMC?†	0.877	-0.028	-0.029	-0.067	-0.028
	(0.328)	(0.047)	(0.04)	(0.059)	(0.022)
Inspectors visited schools?†	0.730	0.078	-0.024	-0.054	0.058
	(0.444)	(0.065)	(0.055)	(0.076)	(0.067)
Number of observations	6,703	285	582	625	612

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

References

Alhassan, RK, Nketiah-Amponsah, E, Spieker, N, Arhinful, DK, De Wit, TFR, 2016. *Assessing the Impact of Community Engagement Interventions on Health Worker Motivation and Experiences with Clients in Primary Health Facilities in Ghana: A Randomized Cluster Trial.* PloS one, 11(7), e0158541.

Ananthpur, K, Malik, K, Rao, V, 2014. *The Anatomy of Failure: An Ethnography of a Randomized Trial to Deepen Democracy in Rural India.* The World Bank, Development Research Group, Poverty and Inequality Team.

Anderson, ML, 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), pp. 1481-1495.

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

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

Banerjee, AV, Duflo, E, Glennerster, R, 2008. *Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System.* Journal of the European Economic Association, 6 (2-3), pp. 487-500.

Barrett, CB, Carter, MR, 2010. *The power and pitfalls of experiments in development economics: Some non-random reflections.* Applied economic perspectives and policy, 32 (4), pp. 515-548.

Beath, A, Fotini, C, Ruben, E, 2013. *Do elected councils improve governance? Experimental evidence on local institutions in Afghanistan.* The World Bank.

Benin, S, Nkonya, E, Okecho, G, Pender, J, Nahdy, S, Mugarura, S, 2007. *Assessing the impact of the National Agricultural Advisory Services (NAADS) in the Uganda rural livelihoods.* IFPRI.

Berman, E, Callen, M, Condra, LN, Downey, M, Ghanik, T, Isaqzadeh, M, 2017. *Community Monitors vs. Leakage: Experimental Evidence from Afghanistan.*

Bernardi, RA, 2006. *Associations between Hofstede's cultural constructs and social desirability response bias.* Journal of Business Ethics, 65 (1), pp. 43-53.

Bernheim, BD, Whinston, M, 1986. *Common Agency*. Econometrica, 54 (4), pp. 923-42.

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

Beuermann, DW, Amelina, M, 2014. *Does Participatory Budgeting Improve Decentralized Public Service Delivery?* Inter-American Development Bank.

Björkman, M, Svensson, J, 2009. *Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda.* The Quarterly Journal of Economics, 124 (2), pp. 735-769.

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

Blundell, R, Dias, MC, 2009. *Alternative approaches to evaluation in empirical microeconomics.* Journal of Human Resources, 44 (3), pp. 565-640.

Bradley, J, Igras, S, 2005. *Improving the quality of child health services: participatory action by providers.* International Journal for Quality in Health Care, 17, pp. 391-399.

Burki, SJ, Perry, G, Dillinger, W, Griffin, C, Gutman, J, Rojas, F, Webb, S, Winkler, D, 1999. *Beyond the center: Decentralizing the state.* The World Bank.

Capuno, JJ, Garcia, MM, 2010. *Can Information about Local Government Performance Induce Civic Participation? Evidence from the Philippines.* The Journal of Development Studies, 46(4), pp. 624-643.

Chung, J, Monroe, GS, 2003. *Exploring social desirability bias.* Journal of Business Ethics, 44 (4), pp. 291-302.

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

Diaz-Cayeros, A, Magaloni, B, Ruiz-Euler, A, 2014. *Traditional Governance, Citizen Engagement, and Local Public Goods: Evidence from Mexico.* World Development, 53, pp. 80-93.

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

Dunning, T, Grossman, G, Humphreys, M, Hyde, S, McIntosh, C, Nellis, G (Eds.), 2019. *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I (Cambridge Studies in Comparative Politics)*. Cambridge: Cambridge University Press.

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

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

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

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

Giné, X, Khalid, S, Mansuri, G, 2018. *The Impact of Social Mobilization on Health Service Delivery and Health Outcomes - Evidence from Rural Pakistan.* World Bank Group, Development Research Group, Finance and Private Sector Development Team & Poverty and Equity Global Practice Group.

Goeree, JK, Yariv, L, 2011. *An Experimental Study of Collective Deliberation.* Econometrica, 79 (3), pp. 893-921.

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

Goncalves, S, 2013. *The Effects of Participatory Budgeting on Municipal Expenditures and Infant Mortality in Brazil.* World Development, 53, pp. 94-110.

Grossman, G, Rodden, J, Tausanovich, Z, Han, A, 2017. *Can Text Messages Improve Local Governance? An Impact Evaluation of The U-Bridge Programme In Uganda.* USAID.

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

Gullo, S, Galavotti, C, Kuhlmann, AS, Msiska, T, Hastings, P, Marti, CN, 2017. *Effects of a social accountability approach, CARE's Community Score Card, on reproductive health-related outcomes in Malawi: A cluster-randomized controlled evaluation.* PloS One, 12.

Heckman, J, Navarro, S, 2004. *Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models.* Review of Economics and Statistics, 86 (I), pp. 30-57.

Hebert, JR, Clemow, L, Pbert, L, Ockene, IS, Ockene, JK, 1995. *Social desirability bias in dietary self-report may compromise the validity of dietary intake measures.* International journal of epidemiology, 24 (2), pp. 389-398.

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

Humphreys, M, De la Sierra, RS, Van der Windt, P, 2014. *Social and economic impacts of Tuungane: Final report on the effects of a community driven reconstruction programme in the Democratic Republic of Congo.* 3ie.

Humphreys, M, Weinstein, JM, 2012. *Policing Politicians: Citizen Empowerment and Political Accountability in Uganda Preliminary Analysis.* Columbia Universities. Unpublished manuscript.

Iacus, SM, King, G, Porro, G, 2012. *Causal inference without balance checking: Coarsened exact matching.* Political analysis, 20 (1), pp. 1-24.

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

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

McKenzie, D, 2012. *Beyond baseline and follow-up: The case for more T in experiments.* Journal of development Economics, 99 (2), pp. 210-221.

Molina, E, 2014. *Can Bottom-Up Institutional Reform Improve Service Delivery?* Inter-American Development Bank.

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

Olken, BA, 2007. *Monitoring corruption: evidence from a field experiment in Indonesia.* Journal of political Economy, 115 (2), pp. 200-249.

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

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

Ram, K, 2013. *Git can facilitate greater reproducibility and increased transparency in science.* Source code for biology and medicine, 8 (1), p. 7.

Raffler, P, Posner, DN, Parkerson, D, 2018. *The weakness of bottomup accountability: Experimental evidence from the Ugandan health sector.* Unpublished manuscript.

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

Steiner, S, 2007. *Decentralisation and poverty: conceptual framework and application to Uganda. Public Administration and Development.* The International Journal of Management Research and Practice, 27 (2), pp.175-185.

Timmons, JF, Garfias, F, 2015. *Revealed Corruption, Taxation, and Fiscal Accountability: Evidence from Brazil*. World Development, 70, pp. 13-27.

Touchton, M, Wampler, B, 2014. *Improving social well-being through new democratic institutions.* Comparative Political Studies, 47 (10), pp. 1,442-1,469.

Uganda Bureau of Statistics, 2017. *The National Population and Housing Census 2014 – National Analytical Report.* Kampala, Uganda.

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

Van Campenhout, B, Spielman, DJ, Lecoutere, E, 2020. *Information and Communication Technologies to Provide Agricultural Advice to Smallholder Farmers: Experimental Evidence from Uganda.* American Journal of Agricultural Economics.

Waddington, H, Sonnenfeld, A, Finetti, J, Gaarder, M, John, D and Stevenson, J, 2019. *Does incorporating participation and accountability improve development outcomes? Meta-analysis and framework synthesis.* 3ie Systematic Review 43. London: International Initiative for Impact Evaluation (3ie).

Zerbe, WJ, Paulhus, DL, 1987. *Socially Desirable Responding in Organizational Behavior: A Reconception.* Academy of Management Journal 12 (2), pp. 250–264.