Social Engineering in the Tropics: Case Study Evidence from East Congo

March 20, 2014

Abstract

Many international interventions try to alter social structures without seeking to change economic fundamentals. The theory is that brief exposure to good institutions leads directly to subsequent adoption. We examine this idea exploiting random assignment of a post-conflict intervention implemented in 1,250 villages in Congo. We measure social outcomes using a cash transfer activity later implemented across all areas and find almost no evidence of effects on governance practices. Unique in its use of an unconstrained behavioral measure and unusual in scale, our study suggests that current conceptualizations of local governance structures, and strategies to alter them, are misguided.

1 Introduction

Since the 1990s, participatory development has become a favored model for delivering international aid. The huge growth of the model reflects two broad trends.[1] First, a conviction that participatory approaches to development yield better results than traditional top-down approaches.[2] Second, that international aid can have a *transformative* effect and yield not just stronger welfare gains but also alter the way political decisions are made at the most local level, rendering local decision making processes more inclusive and more democratic. Community Driven Reconstruction (CDR) projects are quintessential of this approach, representing a model for delivering post-conflict reconstruction aid in a way that not only builds infrastructure but also recreates communities and refashions governance structures. This paper exploits exogenous variation from a large CDR project, in which development aid was made available to 1,250 randomly selected villages in the Democratic Republic of Congo (DRC), to learn whether international actors are indeed able to use aid to transform societies in this way.

A simple argument supports the transformative agenda: institutional innovations that have proved successful in one context will be taken up in others once users are exposed to them and see them operate in practice.[3] That such a transformation might be possible through the introduction of ideas and practices alone resonates with research that emphasizes the role of institutions in economic development (North, 1991; Sokoloff and Engerman, 2000; Acemoglu et al., 2001; La Porta et al., 2008). Much theoretic work on institutions emphasizes the ways in which institutional choice resembles a coordination problem. In some environments, whether one group or another dominates a polity and whether they face challenges or not is a matter of equilibrium selection. If all

[1]Mansuri and Rao (2013) quote a figure of $85bn in World Bank spending in the last decade alone on this broad class of interventions.

[2]e.g. Gandhi (1962); Freire (1970); Scott (1998); White (1999).

[3]Beyond the transformative argument, there are *substantive* and *intrinsic benefits* arguments, that emphasize respectively the improved quality of choice resulting from participation (Mansuri and Rao, 2013) and the gains in autonomy that it can bring (Sen, 2001; Hirschman, 1984)

actors coordinate strategies in one way, the decisions of traditional institutions will be treated as authoritative, but in other available equilibria, such decisions would be subject to challenge (see Young, 2001; Chwe, 2000; Bidner and Francois, 2013). Moreover, in some accounts shifts in expectations may have implications across multiple domains (e.g. Binmore (1998)). Under this logic, small interventions that seek to alter the equilibrium may have long lasting effects: in other words, social engineering on a shoestring may be possible. This idea seems especially plausible for post-conflict development aid, delivered at a time in which war torn societies are commonly characterized as being in flux and lacking institutional strength, even at the most local level (Kaplan, 1994). Indeed in recent years this approach has formed a major pillar of post-conflict interventions in Rwanda, Liberia, Sierra Leone, Afghanistan, Indonesia, the Philippines and elsewhere. In all of these cases, international actors, sometimes working with central governments, have sought to use development aid mechanisms to transform local institutions.

Advocates argue that this development model is effective.[4] But despite the popularity of the CDR model there has been little evidence for the claims made on its behalf (Mansuri and Rao, 2013). The basic assumption behind CDR – that exposure to good governance practices over the course of a few years can alter social behavior – runs largely counter to accounts of the determinants of social behavior that emphasize large structural changes (see Putnam, 1993; Bowles and Gintis, 2004; Nunn, 2008). More recently, there have been a number of studies examining the social and economic effects of these programs. These studies have painted a mixed picture. In their study of a CDR program in Liberia, Fearon et al. (2009) find little or no evidence for economic impacts but some positive evidence for an effect of CDR on the ability of communities to solve some types of collective action problems. Casey et al. (2013) examine a CDR program in Sierra Leone and find evidence of economic effects but no evidence of social effects. In a third study in Afghanistan, Beath et al. (2013)

[4]The World Bank claims, for example, that "CDD operations produce two primary types of results: more and better distributed assets, and stronger, more responsive institutions" (World Bank, 2009).

find some evidence that imported institutions can be effective but only when external groups require that they be employed.

Although ours is a case study, five innovations of the study make it a strong test of the general claim. First, the intervention we examine is one that was expected to have large effects. As part of our design we documented these expectations, gathering data on the priors of researchers and development actors regarding the likely effect of this program. Second, the outcomes we measure provide a direct test; they produce evidence of change on exactly those behaviors that the international aid sought to affect. Specifically, to generate measures of behavioral change we introduced an entirely new unconditional cash transfer scheme ("RAPID"), implemented by local universities and targeted at a randomly selected set of 560 villages – half having been part of the CDR treatment in the preceding three years, and the other half not. These RAPID communities received block grants of $1,000 which they could manage as they saw fit with no oversight and minimal guidance. Given this intervention, a test of the behavioral effects of the CDR model was: did areas that took part in the program engage differently with RAPID relative to those that did not? In employing this approach we forgo some of the control enjoyed in studies that employ individual level behavioral games (such as Avdeenko and Gilligan (2014)) or that use more circumscribed community level behavioral exercises (such as in Fearon et al. (2009) or Casey et al. (2013)). We gain in return a very close mapping of outcomes to the major goals of this type of aid and a strong test of the claims underpinning this approach to development. Third, we enjoy a high level of statistical power. Our study, employing data from over 800 villages in over 400 clusters, is one of the largest randomized trials of its kind. The scale provides unusually strong statistical power and reduces the likelihood of false negatives.[5] Fourth, we use a strategy to help ensure that our results are free from reporting and analysis biases. Our analysis joins a small number of studies that have publicly posted and subsequently followed a detailed pre-analysis plan, specifying not just our core hypotheses but also

[5]This compares to 83 villages in Fearon et al. (2009) and 236 villages in Casey et al. (2013).

the details of all the major analyses and tests we intended to run (see section 3.5).[6] Finally, our study provides one of the first applications of a design-based strategy for assessing spillover effects in this field. Unless segregation patterns are sufficiently strong (Sinclair et al., 2012), spillover effects present a serious challenge to experimental and observational assessment of causal effects. In the past, model-based strategies have been applied (e.g. Miguel and Kremer (2004)) but only more recently have design-based strategies – that are less dependent on assumptions regarding effect homogeneity – been elaborated (see Chen et al. (2010) and Aronow and Samii (2013)).

Our results are overwhelmingly negative. We measure the quality of institutions across five dimensions (participation, accountability, transparency, efficiency, and equity-capture), and find almost no evidence that exposure to development aid alters local political decision making. We believe that these overall negative results are particularly striking in light of the strong statistical power, the strength of the measures, and the transparency of the approach to analysis. Coupled with emerging results from related studies this work calls for a rethinking of the ways that international development actors are engaging in aid, and the scope for inducing behavioral change through the exportation of institutions, while leaving underlying structural conditions intact.

The remainder of this paper is organized as follows. The next section introduces the setting, the intervention and the hypotheses to be tested. Section 3 discusses our empirical strategy including identification, the measurement and estimation strategy and the pre-registration of design and outcome measures. The results are presented in Section 4. Section 5 shows how these results hold up under a large set of robustness checks. We discuss the implications of our findings for development practice and scholarship on the political economy of development in Section 6.

[6]See Humphreys et al. (2013) for analysis of the scope for selecting tests and measurements based on the results they generate, even after hypotheses have been specified. See Casey et al. (2013) for further discussion of preanalysis plans and a striking illustration of the constraints they can place on the scope for selecting narratives.

2 The Intervention

2.1 Post-Conflict Congo

To assess the effects of development aid on decision making we exploit random assignment of the UK government funded *Tuungane* program: a community driven reconstruction intervention in East Congo. The program's area of operation – South Kivu, Maniema, Tanganyika and Haut Katanga – figured centrally in the violence that has engulfed the country over the last two decades. Located in the east, it was home to the start of the First and Second Congolese Wars (1996-1997 and 1998-2003). The latter, with the direct involvement of eight African nations and 25 armed groups, has been the deadliest war in modern African history.[7] Despite the formal end to the war in July 2003, the project area, and in particular the South Kivu province, continues to be an epicenter of conflict.[8]

The effects of the conflict have been far reaching. Basic infrastructure such as roads, schools, and health facilities is lacking, either due to destruction or a lack of investment. For example, our baseline survey, which was conducted in 2007 and contains information about 21,467 household members (2,906 respondents) drawn randomly from almost 600 villages, indicates that the typical household has to walk 45 minutes to reach drinking water (Ref Omitted). The conflict has also led to social upheaval, at least in the form of massive displacement: a full 61% of household members in the sample were reported as having fled at least once at some point during the period from 1996 to 2007.

Despite continued violence, the DRC began to be classified as "post-conflict" by international actors in recent years (Autesserre, 2010). As a result, attention and funding has been redirected from emergency towards development

[7]It is estimated that between 1998 and 2007 the war and its aftermath had killed 5.4 million people, mostly from disease and starvation (IRC, 2007).

[8]The roots and the dynamics of the Congolese conflict are too complex to be discussed in detail here. For a more complete discussion see: Nest (2011), Autesserre (2010), Vlassenroot and Huggins (2005) and Prunier (2009).

and reconstruction programs. After 33 years of rule by the Mobutu regime, the Congolese administrative system is regularly scored corrupt (Schatzberg, 1997).[9] Thus, grassroots-level interventions that avoid working directly with the state were seen as an efficient vehicle to reach those in need. In addition, with its focus on improving infrastructure and its central role for inclusion of the whole community into decision-making about the use and management of development aid (Mansuri and Rao (2004)), the community driven reconstruction model in particular seemed well-suited for the DRC context.

2.2 The Treatment

As is common with large development projects, the general terms of the *Tuungane* program were set by the donor, here the UK government, but the detailed components and objectives were developed in large part by implementing organizations, in this case the International Rescue Committee (IRC) and CARE International. As stated in the original project description provided by these organizations, the *Tuungane* CDR intervention sought to establish and strengthen local governance committees in order to:

> "improve the understanding and practice of democratic governance, improve citizens' relationships with local government, and improve social cohesion and thereby communities' ability to resolve conflict peacefully"

From these broad goals a set of specific primary hypotheses were formed jointly by the research team and the International Rescue Committee (IRC) in 2007 (Ref Omitted).[10] The primary hypotheses on governance outcomes are shown in Table 3 in the Supplementary Material.

[9] In 2013, Transparency International ranked the DRC the 154th most corrupt public sector out of 177 countries and territories.

[10] A broader set of secondary hypotheses relating to variations in implementation, heterogeneous effects, contextual factors, unintended consequences, behavioral outcomes, and measurement strategies were developed prior to data collection and are described in (Ref Omitted).

The core strategy for achieving improvements in governance outcomes comprised the reorganizing of existing settlements into new quasi-communities, the creation of local development committees and finally social interventions tied to the implementation of development projects that seek to demonstrate the practice of public accountability. We describe each of these elements in turn.

2.2.1 Units

The units of operation for the project were 1,250 Village Development Committee areas (VDCs) (containing 1 - 2 natural villages) and 280 Community Development Committee areas (CDCs) (spanning multiple VDC areas). VDC areas covered populations of approximately 1,300 inhabitants and were nested within CDC areas that covered approximately 6,000 inhabitants. Although the interventions we examine operated at the VDC level, the assignment to treatment took place at the CDC level. As we describe below, we take account of this feature in all analyses.

In each VDC area, the project worked through the central instrument of an elected committee (also called the VDC). These committees were formed through open and public elections and consisted of 10 representatives (2 co-presidents, 2 co-treasurers, 2 co-secretaries, and 4 ordinary members). By design, in about 75% of areas, these committees were required to have one man and one woman elected to each position; in the remaining areas this gender parity constraint was lifted (we assess this feature in other work). This representative body was responsible for overseeing the quality of implementation and for reporting back to the populations; the populations would learn that they could select their leaders democratically, charge them with making decisions, and hold them to account. In addition, CDCs were formed by selecting two members from each of the VDCs, by the VDC representatives, and had a similar role at the CDC area level. The VDC level interventions were implemented prior to the CDC level interventions, involved small grant amounts (valued at $3,000), and focused on governance outcomes. The CDC interventions, involved much larger grants (valued at between $50,000 and $70,000), and focused on economic outcomes primarily. As we describe in Section 3.2 below, we exploit this division of labor between the intervention components and focus on the effects of the VDC interventions, taking measurements prior to the implementation of the CDC projects.

2.2.2 Social Interventions and Projects

A primary component of the social intervention was direct participation by the population. Elections were first implemented to create the VDCs. Next, the VDC members, following consultations with the population, selected how to allocate an envelope of $3,000 for a maximum of two projects, and their selection was then put to the population for an up-down vote. The VDC was also expected to be held accountable by the population. To facilitate this, the committee was tasked with sensitizing populations on "the importance of good leadership, and the meaningful inclusion of women and other vulnerable groups" (IRC, 2012). Moreover, general assemblies were convened by the VDCs (with an average of about 4 per VDC) to justify the use of project funds to populations. The sums of $3,000 were disbursed in tranches of $500 each. The condition to receive each subsequent tranche was to successfully manage the previous one and to get the approval of the population.

A second component was training of the committee members. VDC members received two sets of trainings. First, a three day training on their roles and responsibilities, leadership and good governance, gender and vulnerability and the "Do No Harm" principle. A second one-day training focused on financial management, in particular on the necessity of documentation and the roles and responsibilities of the VDC members to ensure adequate financial management of the $3,000 sub-grant (IRC (2012)).

In total 1,811 VDC level economic projects were implemented. A majority of these were small education projects, such as the construction of classrooms, the rehabilitation of classrooms, and purchase of school furnishing items (benches, tables, chairs). See IRC (2012) for a more complete overview.[11]

[11]Broadly, these projects were implemented to plan. IRC records only 26 instances of

2.3 The Intervention in Context

2.3.1 Scale

The *Tuungane* program was one of the largest programs of its kind, operating for approximately four years in 1,250 war-affected VDCs with a beneficiary population of approximately 1,780,000 people, and covering large parts of four major regions: South Kivu Province, Maniema Province, and Haut Katanga and Tanganyika in Katanga province. The project budget, including the budget for CDC projects, was £30m (USD $46m).[12] The programs were implemented in about 4 years on average, with the first (VDC) phase being implemented in about 2 years. See Figure 3 in the supplementary material for an illustration of the timing of implementation across areas.

We note however that although the aggregate numbers are very large by the standards of development projects in the DRC, the *per capita* investments are small. By IRC estimates, about 0.7% of the population (12,510 of 1,780,000) people were directly involved in VDC member trainings. 1,811 village level projects were implemented at a value of $3,707,624 USD over two years, which corresponds to approximately $1 per person per year. A further $14,354,403 was spent on larger CDC level projects. To put these numbers in perspective, the BRA-KDP program in Aceh had investments targeted at around $20 per capita per year and the Millennium Village initiative targets aid at $120 per capita per year. Interventions in Western countries, such as the US stimulus plan involve per capita investments that are orders of magnitude larger.

In the education sector (which was by far the largest sector) an estimated 420 school rooms were constructed and 1,348 renovated, as part of the VDC

VDCs (2.1% of the total) that were excluded from the project due to mismanagement of project funds; and 12 instances of CDCs (4% of the total) where contracted enterprises that had received advance payments failed to carry out the work.

[12] A second phase of the project (*Tuungane* II), now underway, which has a value of £61m (USD$95m)

projects. With an average of about 50 students per class these investments could improve the educational environments of perhaps 90,000 students per year. While this is a major accomplishment, the investment likely provides direct benefits to less than 5% of the population of the project areas. In the health sector, approximately 160 clinics were built or rehabilitated that, if they service entire villages, could reach over 10% of the population; with 5,000 mosquito nets distributed, there are direct gains to nearly 1% of the population, assuming 3 people per net.

For all of these interventions there are possibilities of external effects both in terms of health, education, and economic activity. For instance, because of transmission, improved health for some can have positive health effects for others in the communities and surrounding communities (see Miguel and Kremer (2004) and the analysis in section 5.4). Nevertheless it bears emphasis that, by design, the economic interventions were small with this stage of the project focusing on the social interventions.

2.3.2 Expectations

Despite the relatively small dollar amounts spent in individual villages, there were great expectations for the social effects of the interventions. As we noted at the outset, strong expectations for interventions of this form are not inconsistent with some theoretical accounts of institutional change even if they are inconsistent with historical work on the development of governance structures.

Did development funders and implementers supporting this project expect that it would produce strong effects? Anecdotal evidence suggests they did.[13] More systematically, however, prior to launching our endline data collection exercise we ran a small survey with the population of regional project implementers and project directors (12 respondents) as well as a (convenience) sample of seven researchers working in East Congo and Rwanda on related issues.

[13] For example the website of the group implementing the intervention quotes staff members as claiming this the program "is exciting because it seeks to understand and rebuild the social fabric of communities. [...] Its a program that starts to rebuild trust, it's a grassroots democratization program."

The survey simply elicited beliefs regarding likely impacts on each of the outcomes in different categories. It was not incentivized. The responses showed systematic variation from item to item—which suggests that respondents were not simple optimists. Two thirds of project implementation respondents reported that they thought it "improbable" that beneficiaries would allocate more time to income generating activities; none thought it very likely that household incomes would increase. Yet all but one thought it possible or very likely that there would be improvements in each of three distinct dimensions of governance outcomes. Half thought it very likely that villages would manage projects in a more transparent and equitable way. Researchers were also optimistic about governance outcomes, though they were more optimistic about effects on participation and considerably more skeptical that traditional leaders would become more accountable (most researchers reported that they would not).

During the design stage we also learned qualitatively about the program funders' expectations related to levels of capture. The UK government funders expected that the implementation of RAPID in control areas would result in high levels of capture of development funds. Indeed this expectation was a rationale for the CDR program in the first place. Concerns were so great that special permissions were sought for development funds to be used for the RAPID program in this way, with a view to measurement, in expectation that standard accounting for fund use would likely not be possible.

Access to this prior information is valuable for the simple reason that it was formed prior to data gathering. If the weakness of the intervention seems obvious after the results are in, our information on priors supports the idea that the lessons may extend nevertheless to cases that are currently believed to be model interventions.

2.3.3 Exposure and Reception among Populations

Tuungane had an impact in the traditional sense that the interventions were highly visible in the region. Survey responses to open ended questions about the interventions are largely approving, with many expressing gratitude and support ("God bless *Tuungane* so that it will continue to help other villages"), some repeat the messages of the program quite faithfully ("*Tuungane* helped us and we are very happy because we have leader transparency and cohesion and can go forward"). Others were critical. For example, one chief argued "*Tuungane* marginalized the chiefs entirely even though we are interested; it only listened to the members of the committee which is a serious thing and even the population does not know the first thing about *Tuungane*", and some were critical of the efforts required to make a development project participatory ("For me the project does not do anything well. I think that the project itself should decide what we should do"). In response to a general "approval" question, 81% report the project was "helpful" and only 2% report it to be harmful. Moreover, large numbers were exposed directly or indirectly to the project. *Tuungane* was known by name to almost two-thirds of the population of the area (71% among men, 59% among women). Those that knew about *Tuungane* generally knew who implemented it, although knowledge about the size of grants was somewhat weaker (76% of committee members reported the correct answer of $3000, 48% of chiefs that had heard of *Tuungane* guessed correctly, and just 22% of the general population guessed correctly). About 30% of the population (36% for men, 23% for women) reported participating in votes as part of the project.

The visibility and appreciation of development aid tells us little however about whether aid has a causal effect on decisionmaking. We turn to address this causal question next.

3 Empirical Strategy

3.1 Assignment to Treatment

The selection of communities into *Tuungane* took place through a series of public lotteries. CDC communities were grouped together geographically into

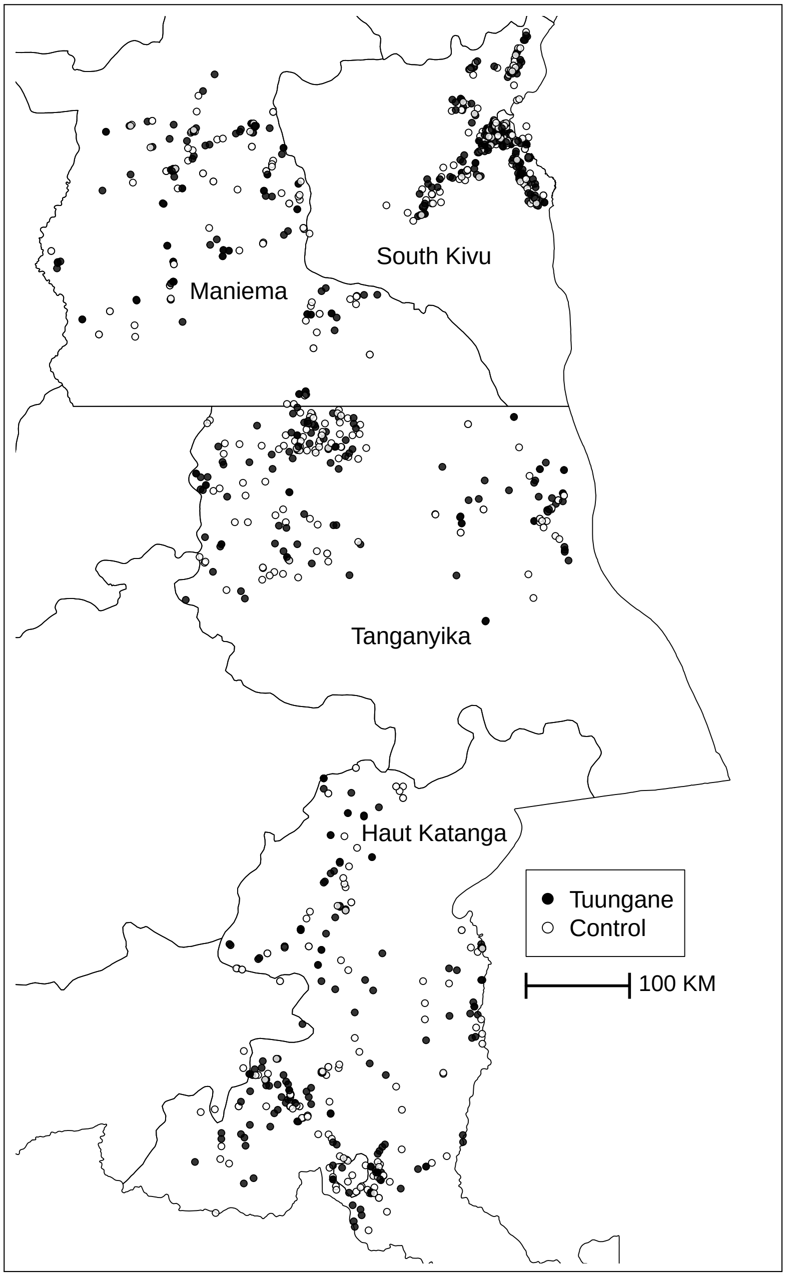
83 “lottery bins” from which CDC project communities were drawn.[14] Representatives from all the potential project communities came together for the lottery, were told briefly about the project, and were able to witness the actual selection of communities (generally done by drawing names out of a hat). In total 600 CDC areas entered lotteries, 280 were selected for treatment and the remaining 320 were in control (see Figure 1).

For the purposes of this study we randomly selected 280 control CDCs at the lottery bin level for study in order to maximize balance in treatment and control units within each lottery bin. These public lotteries have a set of normative advantages, as well as some statistical advantages and limitations. The chief normative advantage is that they provide a limited form of informed consent on the part of communities, both those that benefit from the program and those that do not. Moreover, control communities learn that they could have been a part of the program, and all communities learn that there is a learning component to the interventions. A second, more programmatic advantage is that there is transparency over the selection process and reduces concerns that one community was being unfairly favored over another. A research advantage of selecting communities through a set of lotteries is that within each lottery area there is good geographic balance in terms of the number of treated and control areas, minimizing the chances that treatment communities end up clustered in one area and control communities in another. The flip side of this balance is a somewhat reduced ability to estimate “spillover” effects since clusters of treatment and control villages are contiguous to each other and there is limited variation in geographical distance between treatment and controls. A final concern might be that awareness of the intervention among control communities could lead to jealousy, which could in principle lead to biased results if those communities started performing more strongly or more weakly as a result of not being chosen.[15]

[14] In general, lottery bins corresponded to chiefdoms (“Chefferies”) or sectors (“Secteurs”). For simplicity, we generally use the term chiefdom for both units.

[15] When asked specifically whether the project generated jealousies with other villages, about a third of respondents in *Tuungane* communities answered that it did. In non-*Tuungane* communities about 37% of the general respondents (and 45% of chiefs) that

Figure 1: Distribution of Treatment and Control CDCs



Notes Randomization was implemented at the level of blocks roughly corresponding to chiefdoms and ranging in size from 2 treatment units to 30 units. Source: Authors' drawing.

---

answered this question reported inter-village jealousies. However it bears emphasis that only 15% of the general population and 33% of chiefs in non-*Tuungane* areas had heard of *Tuungane*.

Our survey data allows us to assess the extent to which individuals understood the selection process. We asked a set of survey respondents (that had heard of *Tuungane*) in treatment and control areas how they thought communities were chosen. In treatment areas, 59% of those responding reported that the villages were chosen by chance. Divine intervention was the next most common answer. Few gave traditional explanations such as favoritism by government or NGOs. Patterns in control areas were largely similar in non-*Tuungane* areas although in these areas the vast majority of respondents either had not heard of *Tuungane* or had no explanation for why the project was not implemented in their community.

3.2 What Program Components are Studied?

As discussed in Section 2.2.2, the social interventions took place at the VDC level, while much of the economic investment activity took place subsequent to these social interventions at the CDC level. The research focuses on the first component, the VDC projects, and the timing of the research reflects this focus (see Figure 3 in the supplementary material). There are principled and pragmatic reasons for focusing on the VDC component. The principled reason is that all the major social interventions took place at this level and these components are broadly seen as the key innovative components of the *Tuungane* design, and of community driven reconstruction programs more generally. By implementing research after the implementation of the social interventions, but before the implementation of the major CDC projects, the concerns of a complex treatment (conflating economic and social interventions), though still present, are somewhat mitigated. The practical reason for the focus is that the CDC projects were to be followed almost immediately by a new round of *Tuungane* II interventions, including new social interventions, and it would not have been possible to implement data collection between these rounds of projects.

To the extent possible, we sought to time the research to be at a set interval after the project start in a particular lottery bin (that is, the date of the lottery) and always implemented after the end of VDC projects but before CDC projects came on line. As seen in Figure 3 we were largely successful in this goal.[16] A second implication is that while we are assessing the effects of the social interventions, these social interventions are coupled – as they always are in CDR programs – with economic interventions, albeit very modest ones. Though not our focus here, elsewhere we report the full set of estimated effects on economic outcomes (Ref Omitted). These are in general weak and not significant, which suggests that it is unlikely that the muted effects we describe can be attributed to interference arising from wealth effects.

3.3 Measurement

3.3.1 Tracking the Level at Which Behaviors Might Change

The research seeks to examine social effects of an intervention that operated at multiple levels and so we need to specify the level at which we believe social effects to operate. In practice it is not possible to examine effects at the level of VDC or CDC areas, if only for the practical reason that these units have no meaning in the control areas. More substantively, outside of the context of the *Tuungane* program, these units have no meaning and so looking for effects at this level has unclear external validity. Instead we sought to measure effects at the level of small natural settlements, which in some cases are smaller than the VDC. The principle behind seeking effects at this level, or at levels other than the VDC and CDC, is that to be relevant the program must change expectations and behaviors that matter for outcomes at multiple levels. The evaluation therefore did not seek to measure the effectiveness or persistence of the particular committees introduced, but rather individuals' and communities' practices outside these structures.

[16] Logistic concerns made it impossible to reach this goal exactly (to do this perfectly the data collection would have had to take place over more than 2 years) but the final timing, as shown in Figure 3 (supplementary material), is largely consistent with this goal. The implication for interpretation is that we seek to assess the impact of *Tuungane* three years out from project onset (the median gap between lotteries and the onset of research is 1081 days with a standard deviation of 192 days).

More specifically, for this research we examined behavior of individuals and groups in two natural villages in each CDC area. Surveys were administered in both of these villages, but in addition in one of them we implemented the RAPID behavioral measures, described next.

3.3.2 Behavioral Activities

Since CDR programs seek to affect social outcomes, they confront specific measurement challenges. It can often be difficult to determine behavioral change from survey responses alone. For example, individuals in treatment communities may have learned how to respond in ways that would please outside funders. For this reason, we introduced an entirely new intervention called RAPID ("Recherche-Action sur les Projets d'Impact pour le Développement") to assess behavioral changes due to *Tuungane*. As part of the RAPID process 560 villages were selected to participate in an unconditional cash transfer program in which they would receive grants of $1,000 to be used on projects that benefit the village. Of these, 280 communities had participated in the *Tuungane* program the preceding three years, the remaining 280 had not. Communities were asked to identify and implement projects subject to minimal constraints. The key constraints were that some uses were ruled out if these were likely to result in harm (such as the purchase of weapons) and the grant had to be spent out within a two month period – a somewhat artificial constraint that stemmed from our need to be able to assess the use of funds in a timely manner. There was a moderate encouragement towards distributive projects but these were not required. There was no guidance of any form given as to who should manage the funds and how decisions should be made. The RAPID project was then rolled out in four steps (A-D) over the course of two to three months. The key features are described in Table 1; see also the script provided in Supplementary Materials (D).

The use of an intervention as a measurement strategy gave rise to a number of considerations. One was how best to handle the consent process, given that the intervention was both a real project and a tool for research. For this

Table 1: The RAPID Behavioral Measure

Stage	Description	Features
A	Team A schedules village meeting and conducts surveys	Initial meeting with the village chief to ask him/her to convene a public meeting at which a minimum share of the village population is required to attend. Survey is conducted among 5 randomly selected households.
	Village meeting and additional surveys	The RAPID project is described in a public village meeting. Measures of the quality of participation are taken. The village is asked to take steps towards determining how to use the project funding and identify representatives (with no guidance). The population is informed that at least $900 will be made available. Surveys are conducted with selected groups of those present during the meeting.
B	Collection of forms	Meeting with committee members only. Measures are taken of the village's decisions regarding how to use funding and who is entrusted to manage it. The committee members are informed *in private* that the amount provided to villages will be $1,000 ($100 more than announced to the village), and of the type of audit that will be undertaken.
C	Disbursement of funds by IRC and CARE	$1,000 are disbursed in private to a select group of members identified by the management committee.
D	Auditing	Auditing is undertaken to track the use of all funds, and measure capture, efficiency, transparency, and the accountability mechanisms that were established.
	Follow-up surveys	Surveys are conducted among 10 randomly selected households (5 are those surveyed during Step A). Measures are included to determine the transparency of the RAPID process, the quality of participation in village decision-making, and the efficiency and equity of outcomes.

Notes: Key features of the $1,000 unconditional cash transfer program.

we opted for an approach in which we identified the link with research at the outset. Consent was first sought for the project component, conditional on the ability of audit data to be used for research purposes. Consent was then sought at the village level for gathering more general measures (such as observation of meetings), allowing in principle for villages to accept the project but refuse individual and village level measurement elements. For individual surveys, consent was sought on an individual basis in the usual way. Unusually, to protect villages we agreed in advance that our partners – the IRC, CARE International and DFID – would not receive detailed village-identifiable information on the performance of communities. Moreover for most of the key analyses we relied on behavioral measures rather than attitudinal measures.

3.3.3 Behavioral Measures

Our approach to measure the quality of institutions is comprised of the following five subconcepts: 1. Participation (the extent to which villagers are willing and able to be part of the public decision making), 2. Accountability (the willingness and ability of community members to sanction leaders for poor performance and the willingness of leaders to respond to citizen requests), 3. Transparency (accessibility of information related to public decision making), 4. Efficiency (the extent to which implementation makes good use of resources available), and 5. Equity-Capture (the extent to which benefits of public projects are broadly distributed). These five dimensions capture actions on the supply and demand sides (participation, accountability), the conditions for accountability (transparency), and the quality of outcomes in terms of efficiency and equity. Although corruption is not explicitly listed, corruption in this context may be thought of as a composition of the absence of transparency and equity.

We exploit a variety of approaches to measure these dimensions, employing direct observations by enumerators of behavior in the village, extensive audits in each RAPID village, and large-n survey data from different steps in the RAPID process. In what follows we give an overview of the twelve core families of behavioral measure; we elaborate on these measures in the results section and we provide a complete listing of measures and instruments used to generate them in (Ref Omitted).

1. [Participation] Attendance and Dynamics at Meetings. Observations of behavior in general assemblies provided direct measures of who showed up, who took part in discussions, and which voices dominated.

2. [Participation] Village strategies for Selecting Representatives. Information on behind the scenes the decision making processes —specifically whether voting mechanisms were used to select committee members and projects — were based on triangulated responses from multiple interviews with separated groups conducted at each site.

3. [Participation] Committee Composition. Measures of how communities selected to represent them on the project (for example, how many men, how many people altogether) could be gathered directly from the forms completed by every community.

4. [Accountability] Presence of Accountability Mechanisms. We assessed whether external or internal accountability mechanisms were put in place to oversee villages using independent interviews with committee members (structured with some in a group and some in private) and from a sample of non members in each village.

5. [Accountability] Complaints. We used a survey based measure to assess each respondent's willingness to complain, asking them to indicate whether or not they agreed with a set of thirteen complaint statements.

6. [Efficiency] Quality of Accounting. During the transfer of project funds (Step C), an accounting form was distributed on which the RAPID committee is expected to indicate the total amount made available for the project (out of $1000), and to keep track of expenditures made. We use the presence of this form at the end of the project as an indicator of efficient project implementation.

7. [Efficiency] Information Transmission. We measure the ease of information flow in villages using an exercise in which we provide, during Step A, a random sample of five villagers in 412 RAPID communities with public health information on hygiene and diarrhea. In Step D we re-visit the RAPID communities, but also an additional 396 randomly selected villages. In both types of villages we ask a new random sample of five villagers a set of questions related to public health. Comparing the answers given by the villagers in RAPID communities (villagers that had not received the information during Step A, but who are in the same village as those that did) with the ones given by those in non-RAPID villages (villages where nobody received the health information), allows us to assess the rate of information flow.

8. [Transparency] Knowledge of Project Amount. At the public meeting in Step A the enumerator teams inform the community in a public meeting that $900 or more will be made available through the RAPID project. In fact, during Step B two weeks later a total of $1000 is transferred *in private* to the RAPID committee. Two months later during Step D – as a measure of transparency – we ask a random sample of villagers about the amount of the RAPID grant. This strategy lets us learn about the extent to which basic information on RAPID project finances is shared beyond the committee.

9. [Transparency] Willingness to Seek Information. We also gather a behavioral measure of the the willingness of randomly selected villagers to obtain relevant information about the management of public resources for which they are beneficiaries. At the end of the household survey in Step D, respondents are presented with the opportunity to seek information about the revenues of the last period for either the main school attended by this village or the main health center (the precise units are identified by our teams at each site). They are offered $1 as compensation for attempting to retrieve the information, and an additional dollar upon success. The share willing to take on the task provides a second measure of transparency.

10. [Capture] Financial Irregularities. We measures financial irregularities by assessing the total value of expenditures as assessed by our audit team. For example, to verify measures of price and quantity as listed in the accounting form the auditors visited nearby markets.

11. [Capture] Distribution of Benefits. For communities that chose to use RAPID funds for direct distribution of small assets or consumption goods we gather a direct measures of who benefits by how much. Auditors generated a census of all reported beneficiaries of the RAPID project distributions; a random sample of these was then selected and the auditors verify whether these beneficiaries actually exist, and if so estimate the value of what these beneficiaries received from the RAPID project. Conditioning on villages that selected distribution projects, the standard deviation of benefits is used as a measure of the inequality of benefits distributed within these villages. We note that this measure exists only for communities that selected distribution of some form, a category which is post-treatment.

12. Dominance of Preferences. Our key measure of capture is the extent to which actual decisions reflect the preferences of different types of villagers. We focus on the dominance of the village chief's preferences over preferences of a random sample of other villagers. We produce a measure of chiefly power by comparing the stated preferred project realization by the chief in a private meeting during Step A with the actual project realization. We then compare the predictive power of this measure to a similar one for random villagers.

Table 2 lists the twelve sets of measures along with corresponding hypotheses 2007.

3.4 Estimation

Because of the random assignment to treatment, treatment and control units within each lottery bin are balanced on observable and unobservable characteristics in expectation (for more information on balance in realization see Supplementary Material (C)). As a result, subject to conditions specified by Rubin (1974) and others, comparing mean outcomes in *Tuungane* communities with those in control communities gives unbiased estimates of the causal effect of the intervention. Corrections are needed however to account for small differences across lottery bins. At the time of randomization, targets (number of CDCs to be selected for the program) were set for each lottery bin and in general these targets were close to 50%. Nevertheless the exact targets vary between bins, sometimes because of integer problems (in some 3-village bins, just one village was selected, in others two) and sometimes because of the programmatic needs to have larger numbers of treated CDCs in different regions. The result is that not every unit has the same propensity to enter the program; that is, units in different bins were selected with different probabilities (but units in a given bin were selected with the same probabilities). Thus comparing raw outcomes in treatment and control CDCs would produce a biased estimate of the effect of treatment, since treatment CDCs for bins where many communities were selected into treatment would be over-represented, distorting the comparison of outcomes. We take account of this fact by applying inverse propensity score weights to every unit (the inverse of the share of units from each lottery bin that were targeted for treatment or control).[17]

For some analyses we have access to multiple, related measures. In this case distinct issues of interpretation may arise. For example, it may be that all measures trend positive, but none is individually statistically significant. In such a case it is possible that effects are jointly significant across the family of measures. Conversely, it may be that by chance a certain measure is significant in a family while most are not, or even trend in the wrong direction. In such cases it is possible that there are no significant effects across the family of measures. For these analyses, in order to generate a meaningful summary of multiple effects within each family, we follow the approach of Kling et al. (2007) and create standardized indices of outcomes on related items.[18] This practice has also been adopted in Casey et al. (2013), Fearon et al. (2009) and Beath et al. (2011), among others.

[17] If p_j is the probability of being assigned to treatment then the inverse propensity weight is $1/p_j$ for treatment units and $1/(1-p_j)$ for control units. In practice targets were set so that there were often more control units than treated units and so, to maximize efficiency in data collection, we undersampled from control CDCs and modified weights accordingly. Thus if for example there were 3 units in a bin and one was assigned to treatment, the propensity score weights would be 3 on the treated unit and 3/2 on the control units. If in practice we randomly selected only one control unit for research then the weights on control units would be $2 \times 3/2 = 3$, resulting in identical weights for the one treatment and one control unit in our sample.

[18] This is done as follows. First we redefine each of the variables of interest in a family, so that higher values for each variable imply positive effects. Second we rescale each of the redefined variables using the (weighted) mean and standard deviation of the control group units. The index is then the standardized average of the redefined rescaled variables. For these measures the outcome in the control group is 0 by definition, and effects of the CDR program are measured as units of a standard deviation of control areas. Loosely that means that if an effect of "1" is observed then the average difference between treatment communities and a control communities is as big as the average difference between any two units in the control group. There are many factors that generate the standard deviation of outcomes between communities in the control group; if the treatment is able to increase outcomes of treated areas on average by the standard deviation of control groups, then the treatment plays alone a very large role in affecting the outcomes of communities that would otherwise not have been treated. On this scale a treatment effect of .2 or .4 would be a large effect.

3.5 Registration and Mock Report

Conscious of concerns that empirical analyses can suffer from a propensity to favor reporting "significant" findings in classical statistical tests, and that this practice can lead to bias in assessment of effects, we sought to employ a form of pre-registration of our research design. Pre-registration has been promoted in medical fields and is starting to be used in social sciences (De Angelis et al., 2005; Casey et al., 2013; Humphreys et al., 2013). In practice, all of our analysis were based on hypotheses that were developed *ex ante* (in 2007) and specified without reference to evidence on treatment effects. Perhaps more critically, the core analysis was developed and coded by the research team at a time when less than 5% of data was available and without reference to actual outcomes. Instead simulated data was analyzed and the results were written up in a "mock report" – a complete report with analysis and discussion of results– and posted online.[19] The analysis presented here differs from the mock report in four ways.

First, we focus here on the behavioral tests of governance effects. All other results are provided in supplementary material (see (Ref Omitted)).

Second, we implement an additional test of the effect of *Tuungane* on the propensity to complain conditional on funds missing. In our analysis plan we sought to estimate the marginal effect of *Tuungane*, controlling for funds missing. A more natural test is to assess whether *Tuungane* increases the interactive effect: does *Tuungane* increase the responsiveness of complaints to funds missing. This approach we feel is more faithful to the hypotheses being examined, however we note that it is less well identified. Below we provide both results, and note that while no treatment effect was found for the unmodified test, there is evidence for an effect under the revised approach.

Third, the index on health information flows is changed to focus only on items that were provided to peers (excluding items provided exclusively to chiefs). This reflects the intention of the original measure but produces no

[19] This experiment with comprehensive registration, in which there is full detailing of an analysis and reporting plan, revealed both benefits and practical difficulties.

substantive effect on results.

Finally, for clusters of measures we implement summary analyses, using mean effects indices, as described in Section 3.4. These result, reported here, make for easier interpretation of the multiple results described in given tables (disaggregated results are available in (Ref Omitted).

4 Results

Did the CDR program induce institutional change? That is, did the *Tuungane* project, by bringing decisions down to the local level in addition to material support, transform the nature of governance itself?

We find no evidence that it did. Table 2 provides the results categorized by the five sub-components of interest: participation, accountability, efficiency, transparency and capture. We describe the estimated level for each measure in control communities. This is interpreted as the expected outcome in the absence of the program. We provide the estimated effect of *Tuungane* in the subsequent column.[20]

Participation. Did *Tuungane* increase the extent to which villagers are willing and able to be part of public decision making? The first meeting in Step A provided the opportunity for communities to learn more about the RAPID project and discuss what they would like to do with RAPID funding. The first row in Table 2 shows that in control communities on average 130 adults participated in the public meeting. In *Tuungane* communities, compared to this control group, on average two fewer individuals showed up; a difference that is not statistically significant. The next row provides results for the extent to which individuals took part in public deliberations. We find that, on average, fourteen interventions are made, with only marginally fewer interventions in *Tuungane* communities. Exploring the type of intervenor, we

[20]Given the hypotheses of the program, these tests are conducted as "one-sided tests" – we are thus interested in testing whether there is sufficient evidence to reject the hypothesis that the program did not have any positive effect.

find that in control communities men and elderly dominate the discussion, being responsible for 71% and 55% of the interventions, respectively, while the chief is only responsible for 3% of the interventions. The third row in Table 2 shows that *Tuungane* had no impact on the dominance of men in public decision making. We find similar non-results for the dominance by the elderly and the chief (not reported).

Between step A and B, RAPID communities were required to select both a committee and a project as part of the terms of receiving funds. After leading two simultaneous focus groups, one with members of the committee and a second with ordinary villagers during step B of the RAPID process, our enumerator teams coded the selection process as either electoral, through lottery, by consensus, imposed by the chief or elders, other or unknown. Approximately 43% of committees and 31% of projects were coded as selected through election, and 71% of committees and 73% of the projects were selected through either election, lottery or consensus. As the composite measure, combining these four indicators, in Table 2 illustrates, we find no evidence that participation in *Tuungane* leads to greater adoption of participatory processes in the selection of the committee or projects.

A final measure of participation is the composition of the RAPID committee. There was no constraint placed on the composition of these committees other than size (at least 2 members and no more than 8). Our composite measure includes the number of women, the number of men, the total size, and the share of women on the committee. We find a strong tendency towards male domination of committees: of 452 committees, 28 had gender parity, two had more women than men, and the rest had more men than women. On average about 1 committee member in 7 was a woman (18% in control; 20% in treatment). Again, on the composite index we find no statistically significant difference between *Tuungane* treatment and control communities.[21]

Accountability. Did *Tuungane* increase the willingness and ability of com-

[21]Looking at the number of women and the share of women individually, we do find evidence that the *Tuungane* program had an impact though significance is lost when we examine the index.

Table 2: Main Results

Concept	Measure (Measure Number)	Comp	Hypothesis	Control	Effect	(se)	N
Participation	Meeting Attendance (1)	No	H2, H3	130.48	-1.98	(7.40)	455
	Interventions in Meeting (1)	No	H2, H3	14.21	-0.49	(0.42)	457
	Dominance in Discussion (1)	No	H2, H3	70.81	0.52	(1.49)	457
	Participatory Selection Methods (2)	Yes	H2, H3, H6	0	0.07	(0.09)	451
	Committee Composition (3)	Yes	H2, H3	0	0.08	(0.10)	452
Accountability	Presence of Accountability Mechanisms (4)	Yes	H5	0	0.00	(0.10)	414
	Private Complaints (5)	Yes	H5	0	0.02	(0.07)	3647
	Private Complaints | Management (5)	Yes	H5	0.29*	0.68**	(0.38)	3502
Efficiency	Existence and Quality of Accounting (6)	Yes	H1	0	0.01	(0.11)	399
	Information Transmission (7)	No	H1	9.66***	-1.41	(1.56)	3800
Transparency	Knowledge of Project Amount (8)	No	H4	38.60	1.52	(3.21)	3685
	Willingness to Seek Information (9)	No	H4	37.70	3.84	(3.28)	1406
Capture	(Fewer) Financial Irregularities (10)	No	H2	851.51	3.52	(20.74)	394
	Number of beneficiaries (11)	No	H2	40.95	3.28	(5.52)	154
	Inequality of benefits (11)	No	H2	8.59	0.56	(1.52)	127
	Dominance of Preferences (12)	No	H2	0.04***	-0.01	(0.03)	2666

Notes: For a more complete discussion on each measure, see (Ref Omitted). One, two or three asterisks indicate, respectively, significance levels at the 10%, 5% and 1% at one-sided tests. Errors are presented in brackets and are clustered at the village level. For the measures "Private Complaints | Management", "Information Transmission" and "Dominance of Preferences" we estimate a model of the form $Y = \beta_0 + \beta_1 X + \beta_2 T + \beta_3 XT$, and report β_1 in the control column and β_3 in the treatment effect column where X corresponds, respectively, to fund mismanagement, an indicator for RAPID, and an indicator of being the chief. All analyses employ propensity score weights and allow for arbitrary clustering of standard errors at the CDC level.

munity members to sanction leaders for poor performance, and the willingness of leaders to respond to citizen requests? We find that in the majority of villages no mechanisms had been put in place to oversee the use of RAPID funding. However, 13% of respondents indicated that an external accountability measure (such as a distinct committee) had been put into place, and another 13% indicated that the committee has been required to report its actions to the community as a whole. As the composite measure in Table 2 indicates, *Tuungane* did not lead to a greater propensity to put accountability mechanisms into place.[22] During Step D, we asked 10 randomly selected

[22] The composite measure includes nine variables: three measures (external accountability measure, committee, or any mechanisms) from three different sources (focus group with the RAPID committee, interview with two RAPID committee members, interview with 10

respondents to indicate whether or not they agreed with thirteen pre-selected complaints. As calculated by an index of the average propensity of villagers to issue complaints, results in Table 2 suggest that levels of complaint are no higher in *Tuungane* areas than in control. Finally, to capture the extent to which these complaints reflect a propensity to complain given that there is something to complain about, we examine the effect of *Tuungane* on the responsiveness of complaints to the level of fund mismanagement. Here, in this one unregistered test, we find the strongest evidence for a *Tuungane* effect, with an approximate tripling of the responsiveness of complaints to fund mismanagement.[23]

Efficiency. Did *Tuungane* communities make better use of the resources provided by RAPID? On average, in 82% of the villages, committees had the accounting form present upon arrival of the audit team during Step D. Approximately 78% of the funds were formally accounted for as calculated by the RAPID Committee (and 83% when calculated by the audit teams). In addition, 56% of the money the committee made available for the RAPID project (of the $1,000) was justified by receipts, and 46% was justified with receipts deemed credible by the auditing team. Table 2 presents the composite index taking these individual measures into account. We do not find evidence of an impact of *Tuungane* on the existence and quality of accounting.

A second behavioral measure of the extent to which the community can function efficiently outside of the RAPID process is generated by examining the extent of effective transmission of information within villages. Section 3.3.3 discussed how we provided public health information on hygiene and diarrhea to a random sample of five villagers during Step A. In Step D we randomly selected five villagers in non-RAPID villages, and five new villagers in RAPID villages. We find that the latter—i.e. those living in villages where we distributed the information to other people—score on average 10 points higher

random villagers).

[23]Note that the estimates of the effect of mismanagement on complaints is not identified for either the control or the *Tuungane* groups; however the difference is identified, under the assumption that quality of implementation is not itself affected by treatment for any units.

on a set of questions related to the public health information we provided. This result indicates that information spreads. As Table 2 shows, *Tuungane* had no impact on this information transmission. The RAPID effect is smaller by 1.4 in *Tuungane* villages than in non-*Tuungane* villages, suggesting that *Tuungane* villages may do marginally worse at information transmission.

Transparency. Did *Tuungane* increase the population's access to information related to public decision making? We implement a particularly strong behavioral measure to learn the extent to which basic information (beyond what we make known to villagers) on RAPID project finances is known in villages. Enumerator teams told communities in a public meeting during Step A that $900 or more will be made available to the village, but in Step B a total of $1,000 is transferred *in private* to the project committee. During Step D we ask 10 randomly-selected villagers in each community about the size of the RAPID grant. Table 2 shows how, on average, 39% of all respondents (and 56% of those respondents that gave an answer) report the correct answer of $1,000. This suggests relatively high levels of transparency. We find no evidence, however, that there is a difference to guess the $1,000 correctly when we compare *Tuungane* with control communities.

We implement another behavioral measure to assess the willingness and ability of randomly selected villagers to obtain relevant information about the management of public resources. Of the 1,406 respondents, approximately 38% of those in control communities were willing to seek information (receiving one dollar for the attempt, and an additional dollar upon success). The people that refused gave various reasons: that it is not appropriate to ask for this information (76), that the respondent did not have time (75), that the game is strange to them (50), that the husband of the respondent refuses or would refuse the collection of this information (13), and other reasons (192). Overall, this suggests broad challenges to accessing basic financial information. We do not find evidence however that taking part in the *Tuungane* intervention makes valuable information about public resources more accessible.

Capture. Did *Tuungane* increase the equitable distribution of project benefits, and decrease the concentration of such benefits among elites or particular subgroups? Our most important measure of capture is the amount of the $1,000 grant that our auditors are unable to account for during their two day community audit. As seen in Table 2, on average $852 of the $1,000 could be verified by the teams. There is no significant relationship between *Tuungane* and traceability of funds.

A second behavioral measure of capture is the extent to which benefits are distributed broadly or narrowly in villages. We find that, on average, around 40% of the households in the villages with projects of private distribution claim to have received private transfers from the RAPID project. There are on average 4% more beneficiaries in *Tuungane* villages, but this difference is not statistically significant. Focusing our attention on the dispersion of the benefits, we calculate a simple standard deviation of the distributions that took place (in dollars) to represent the average difference in the amount received between two randomly selected villagers. On average, in control communities this standard deviation is around $9. Our results indicate that, again, there is no statistically significant difference between *Tuungane* and control communities.

Finally, we provide results from a behavioral measure that captures the extent to which actual decisions reflect the preferences of different sorts of villagers. We find that the project realization (obtained during Step D) coincides better with the stated preferences (taken during Step A before the village meeting) of the chief than those of the villagers. We find that in control areas the chief's prior preferences are 4 percentage points more likely than those of a randomly selected villager to coincide with actual projects. We thus find some evidence of chief dominance but no evidence that the *Tuungane* program reduces the degree of this dominance.

To summarize, barring one measure, we fail to find evidence for an impact of *Tuungane* on any of the major behavioral measures of change.

5 Robustness

There are a number of possible threats to the validity of the null findings we presented in the previous section. We consider six classes of threat and describe evidence from robustness tests that address these concerns.

5.1 Attrition and Missing Responses

A first threat to validity stems from missing responses. The study was designed to gather survey data in a sample of 1,120 villages, half of which were selected for the RAPID project. Different targets were set for different items but the most common data (the household survey) was to be gathered for 10 households in RAPID villages and 5 households in survey-only villages. Given that there were 560 RAPID villages and 560 non RAPID villages this makes a total of 8,400 households (for some items gathered only in RAPID or only in survey-only areas, the targets were 2,800). However, the survey teams successfully collected final (Step D) data on 72% of villages and 62% of individuals, with higher numbers gathered for steps A and B. The full complement of targeted data was not gathered for a number of reasons.

The most significant site of missing data is Maniema province. Political tensions in the run up to the November 2012 presidential elections led to the expulsion of the Maniema teams shortly after the launch of Step D. This led to the loss of 89% of RAPID villages and 89% of survey-only villages for all measures based on Step D, or involving a combination of steps in this region (the data loss was greater for Step D than for Step A and Step B data, which were more advanced at the time of the expulsion).[24] This loss covered entire lottery bin areas, affecting treatment and control units alike. While it affects the range of areas to which our results can speak, as well as our statistical power, we do not think that this loss is plausibly related to the treatment status of units and is thus unlikely to induce bias.

[24] A total of 62/147 RAPID villages received Step A, a total 7/147 RAPID villages received Step D. The same number of survey-only villages received Step D.

A second significant source of missing data is the inaccessibility of some regions for safety and security reasons. Such losses account for 36 village losses outside Maniema, with balance between RAPID and survey-only villages. However, since these also affect clusters of regions containing both treatment and control areas in nearly equal amounts, they are not plausibly related to treatment status.

A third reason for data missingness is failures in the field, ranging from loss, damage, or theft of PDAs, water damage to paper surveys, or enumerator error in the implementation of surveys or particular questions. Given the difficulty of the environment in the DRC, this third category is relatively small affecting a total of 7% of surveys in surveyed villages. This loss is statistically unrelated to treatment status. The fourth area of data loss is due to non-response on particular questions by subjects, again here we have not found evidence that missingness is associated with treatment status.

A final concern is survey non-response. An examination of household survey data suggests that there was no response from 2,200 out of the 5,473 of the initial households selected for the endline survey; these were replaced by neighboring houses. The major reasons for nonresponse were absence of an individual of the indicated gender (712), empty households (617), refusal for any member of the household to be interviewed (95, or <2%), and households that were not found by the survey teams (360).[25] These non-responding households were split almost exactly half and half between treatment and control units suggesting that household missingness is not correlated with treatment. The implication of this is that household level results can only be interpreted as the attitudes of individuals in households accessible at time of survey.[26]

[25]In Step A enumerator teams created a sampling frame of all households in the village. From this ten households were randomly selected: five to be interviewed during Step A and Step D, and five to be interviewed only during Step D. For non-RAPID villages the sampling frame was created, and five households selected, during Step D.

[26]And more precisely of accessible individuals in accessible households.

5.2 Noncompliance

A second threat to the validity of the interpretations offered here is treatment noncompliance in the sense that areas that were selected by lottery to participate in *Tuungane* did not, and areas that were not selected did in fact participate. Survey data indicates that approximately one in seven chiefs either deny that *Tuungane* took place in a *Tuungane* community, or claimed that it did take place when according to records it did not. For all cases with discrepancies between our data and chief reports we asked the IRC to confirm whether the project did or did not take place in these areas. IRC records of where *Tuungane* did take place matched our records of where *Tuungane* ought to have taken place in 77% of these ambiguous cases. This suggests that the discrepancy is due either to weak impact, poor recall by chiefs, or enumeration error. The check left 51 cases out of 806 of possible noncompliance and/or database error. For the analysis in this paper we use our database measure of units selected by lottery which, assuming our database is correct, can be interpreted as "intent to treat" effects (albeit with a high compliance rate). In a robustness test we analyze results under the assumptions that our database is incorrect, that the IRC data is correct, and there is no failure of compliance. Our results (see Table 5) are similar.

5.3 Treatment Heterogeneity

As seen in Figure 3 there is heterogeneity both in the timing and length of project implementation and the timing and length of data collection relative to project implementation. Broadly the research schedule sought to follow the timing of the start date of implementation of *Tuungane* in each area, although the research schedule was more compressed. While the timing of project initiation spanned approximately two years (with the first lottery date being in July 2007 and the last in April 2009), the data gathering spanned approximately one year (with the first village that was visited with step A of RAPID in October 2010 and the last villages visited for step A in October

2011). Thus, in general, and by design of the research, areas that launched late also had a shorter lag between start and measurement. The median gap was 1,185 days, and 90% of cases had a gap between 871 and 1,202 days. These timing decisions however all took place at the level of lottery bins, all units in lottery bin areas were first exposed to the project at the same time (although projects started at different times) and were visited by the research team at the same time, thus ensuring strong balance in timing issues between treatment and control areas at the bin level. The implication of this heterogeneity is that the results should be seen as the average of a set of experiments that varied in time to measurement.

5.4 Spillover effects

In principle a development intervention like *Tuungane* could produce positive spillover effects across communities. Such spillovers would imply greater program effectiveness but could lead to smaller estimates of treatment effects. Three features however suggest that this is not likely. First, assuming indirect effects are weaker than direct effects, strong treatment effects would result in smaller but still positive estimated effects. Second, in this case "communities" are comprised of clusters of villages meaning that most treated villages are surrounded by treated villages and most control villages by control villages, which limits the scope for spillovers to control areas. Third, as discussed in Section 3.1, populations in control areas report very low levels of knowledge about *Tuungane*.

Nevertheless, bias due to spillovers remains possible. Moreover, if spillovers are in operation there is an independent interest in assessing the magnitudes of these spillovers. We use a design based strategy to assess the presence of spillover effects.[27] Specifically, we define an "x-km indirect effect" as the effect of being within x kilometers of a *Tuungane* village that is member of another

[27]See e.g. chapter 8 in Gerber and Green (2012) for more details. Another, more basic approach that uses the distance to the nearest *Tuungane* village as an alternative treatment (conditional on lottery bin and shortest distance to any village) in order to capture spillover effects, produces similar results.

CDC. [28]

The propensity of being exposed to such a treatment effect depends not just on the random assignment of units to treatment but also on the location of any given unit with respect to others. Fortunately we can make use the random assignment to recover these propensities since they are determined, albeit in a complex way, by our original randomization. To calculate these propensities we randomly re-allocate the *Tuungane* treatment to obtain 5,000 possible assignments of all units to treatment and control, employing the same scheme as used in the original randomization. We then assess for each unit, the probability of receiving direct treatment, indirect treatment, and each combination of these. To avoid instability arising from large weights we limit the analysis to villages that have at least a 10% to 90% probability of being in each of these groups for any value of x.

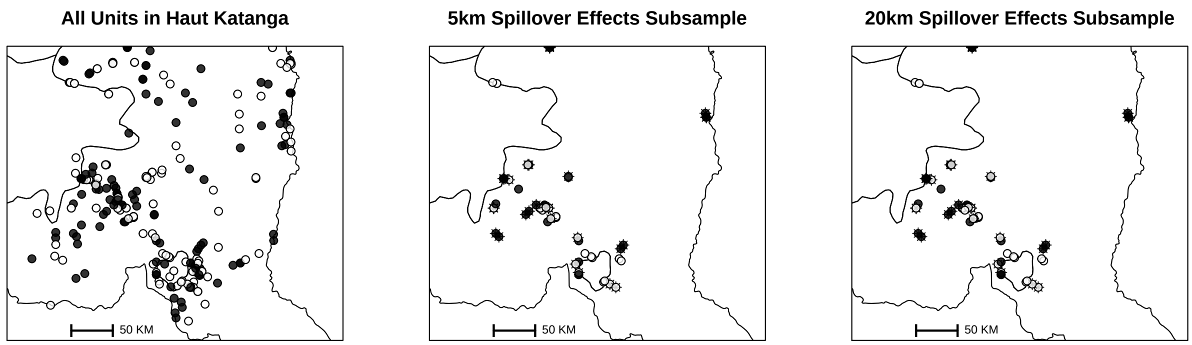
Estimates of treatment effects are then generated by comparing outcomes in each combination of conditions with inverse propensity score weighting using the *known* propensity for each unit of being in each condition. We test the sharp null of no effects using a randomization inference procedure (Fisher, 1935).[29] That is, for each of the 5,000 re-assignments to *Tuungane* we calculate the estimated effect of each treatment type for each outcome of interest. Combined, these estimated effects produce a reference distribution under the sharp null. We compare the actual estimated effect to this distribution and estimate how likely it is we would have obtained results as strong or stronger than our estimated effects under the sharp null.

We conduct our analysis for both a 5km radius spillover treatment and a 20km radius spillover treatment. We highlight (and illustrate, in Figure 2) that when we examine different conceptualizations of the treatment effect we *simultaneously alter our samples*. The intuition is that a unit in a block

[28] Note that for the spillover analysis missing data affects both the set of units in the study *but also the measures of exposure to spillovers*. Our results assess the effects of being close to a treatment village for which we have locational data. We ignore this distinction in light of the small number of units with missing data (we have GPS locations for a total of 1,020 of the 1,120 villages).

[29] See also: Barrios et al. (2012), Small et al. (2008), and Ho and Imai (2006).

Figure 2: Population for Assessment of Spillover Effects



Notes First panel shows the distribution of all treatment and control villages in a section of Haut Katanga. The middle panel shows the subsample of villages that had moderate $(0.01-0.9)$ propensities of being exposed to direct *and* indirect effects of treatment, using a 5km rule for indirect exposure. Indirectly treated units are marked with a cross (and these may themselves be directly treated or not). The right panel shows the corresponding subset for a 20km rule. Note that here changing the definition of the spillover treatment *changes the subset of cases* that have a non-extreme propensity of being exposed to spillovers. Source: Authors' drawing.

with many units, but that has no neighbor within a 10km radius, has a 50% chance of receiving the direct treatment but a 0% chance of being exposed to the indirect treatment of "having a treated neighbor within 5km." Such a relatively isolated unit would drop out of our analysis of a 5km treatment effect. The same unit however might be retained for an assessment of the effect of being within 20km of a treatment village. Villages in more clustered areas may enter the analysis set for the first analysis but not the second (since these may have a 100% chance of being indirectly treated under the first definition). In fact, analysis for the 5km (20km) radius retains 109 (199) units, with only 20 villages being in both groups.[30] Finally, we note that while our estimates

[30]Setting x to 5 yields 516 (504) villages that are (not) directly treated, and among those 450 (570) villages that are (not) indirectly treated. Setting x to 20 yields 504 (516) villages that are (not) directly treated, and among those 874 (146) villages are (not) indirectly treated. Conditioning on these villages having a 10% to 90% probability to be in each combination retains a total of 109 villages: 19 neither direct nor indirect, 35 not direct but indirect, 30 direct and not indirect, and 25 direct and indirect. At a 5km radius these categories total, respectively, 44, 55, 47 and 53 (summing up to 199).

of spillover effects depend on the assumption that in each analysis we have correctly modeled the structure of spillovers, our test of the sharp null does not (Bowers and Fredrickson, 2013).

Our results (see Table 5) provide little evidence for direct effects once we take account of spillovers, or for spillover effects. There is somewhat stronger evidence for gains on our measure of committee composition from treatment once we take account of spillovers in a 20km radius but no evidence of indirect (spillover) effects on any measure.

5.5 Desirability Bias

Because many CDR programs seek to affect social outcomes, a major concern for assessments is that results may be contaminated by social desirability bias. Individuals in treatment communities may have learned how to respond in ways that would please outside funders. This concern might be particularly pertinent for *Tuungane* because the measurement took place while the CDC stage of *Tuungane* was still being implemented – this is illustrated in Figure 3. Thus insofar as respondents felt that the measurement was associated with the program, this may result in a social desirability bias in responses. To minimize this risk we took efforts to ensure that the research was not associated with the IRC or CARE International.[31]

We also introduced a survey variation to investigate the social desirability bias directly. We asked a randomly selected set of respondents the following question: "Do you agree with the idea that elections are the best way to choose

[31] More specifically, the teams introduced themselves to the villages as affiliated with the Official University of Bukavu (in Maniema and South Kivu) or the University of Lubumbashi (Haut Katanga and Tanganyika) and that the project RAPID was implemented by their respective universities in cooperation with Columbia University in New York City and was funded by the British government. Although we sought to minimize any connection with IRC and CARE we also adopted a policy of no deception: if respondents asked directly about IRC or CARE involvement, team members acknowledged their involvement, emphasizing their role in disbursing funds. Moreover, the IRC and CARE International employees that visited villages to distribute the project funds during Step C were assigned to areas in which they had not worked previously so that they would not be identified as staff by populations.

community representatives for positions with technical responsibilities?" For one randomly selected subgroup the question was preceded by the statement "Many NGOs in the region think that election are not the best way to choose community representatives when it comes to an appointment with technical responsibilities"; another subgroup was told "Many NGOs in the region are of the opinion that elections are always the best way to choose community representatives for technical posts." Comparison of answers allows us to assess the degree to which respondents seek to provide answers that they think NGOs want to hear. The results are provided in Table 6 in the supplementary material. We find very strong evidence for a social desirability bias: individuals respond in ways that they think would please outside funders. Importantly, however, we do not find evidence that this bias is affected by exposure to *Tuungane.*

5.6 Specification

Finally, out of concern that analysis decisions resulted in false negatives, we also undertake a series of robustness tests to examine the extent to which the non results are sensitive to various features of our specification. First, we estimate all effects at the village level, where the variables are aggregated using individual sampling weights. The village level analysis is then done using propensity weights only, limiting the extent to which extreme sampling weights can influence cross village comparisons. Second, we control for lottery bin fixed effects. Finally, we generate results (at the village level) using propensity weights adjusted to assess VDC level sample average treatment effects rather than population average treatment effects. These weights have lower variance and may provide more precise estimates. Our results (see Table 5) are robust to these different specifications.

6 Conclusion

We exploit exogenous variation from an intervention in which development aid was made available to 1,250 randomly selected villages in East Congo. The aid program was predicated on the idea that governance problems were rife at the local level, but that these problems could be addressed by marshaling a type of demonstration effect: exposure to practices of democratic and accountable decision-making would lead to the broader adoption of these practices. Optimistic as this simple idea sounds, it underlies a large class of development aid projects including many of the largest interventions in post-conflict areas, but also a much wider set of interventions supported by bilaterals, the World Bank, and other international organizations (Mansuri and Rao, 2013). This type of logic is common to many interventions in the areas of development and governance that seek to export institutions while leaving structural features, such as income distribution or property rights, intact.

Our assessment of this idea suggests that it is not right. We find no evidence to support the claim that exposure to major governance interventions altered local political decision making. In fact, on all measures bar one, we fail to find support for an impact of the CDR project across all five dimensions of institutional quality used in this study (participation, accountability, transparency, efficiency, and equity-capture).

Our conclusions, especially if they find support in future studies, have implications both for scholarly work and aid practice.

First, there are implications for scholarship on local governance structures in developing societies. Development actors frequently adopt a chiefs-as-despots model, and their view is confirmed in some academic accounts of local authority (Acemoglu et al., 2014; Murphy, 1990; Mamdani, 1996). Yet, at the same time there is recent empirical evidence pointing out that chiefs command the respect of rural people (Logan (2013)) and that they might be particularly good project managers (Turley et al. (2014)). If a single pattern stands out across our measures it is the relatively high level of general participation, public information, and equity in decision making across these societies, suggesting that the despotic model is inaccurate. Whereas funders of the RAPID project worried that large shares of funds would be diverted in both treatment and control areas, in practice diversion was low in both. Our results suggest that the despotic/representative dichotomy may be overdrawn and that future work would do better to assess the conditions under which local chiefly structures are more or less accountable.

The generally good performance on governance indicators does not imply however that social structures already operated everywhere as international actors wanted to see them. The interventions sought explicitly to increase the role of women in social decision making for example, mandating in most cases that 50% of CDV representatives be women. In the unconstrained RAPID program however only 20% of selected representatives were women in treatment areas (18% in control areas). Less than half the communities used any kind of election to select committees or projects. The lack of change on these fronts suggest that existing institutional practices are resilient. We cannot determine from this analysis whether, given the distribution of resources and interests, different equilibrium institutions exist, but our results do suggest that given current endowments, shifts of the form sought by development actors are not so easily induced.

Second, there are implications for aid practice. The most obvious implication is that if the purpose of international interventions is to alter local governance structures, the current approach may be flawed. What the flaw is depends on the reason for failure, which is not well identified by this study. One possible reason for failure is that the investments were too small – though we note there are no results in the literature on dose-response for this type of intervention. Another is that the process, like many others, sought to bypass pre-existing institutions and so failed to engage with the relevant centers of power. When we examine our data we find the intended agents of change that were trained by the intervention were not the actors that societies turned to manage collective action problems, a feature consistent with this view. Patterns found by Beath et al. (2013) also support this view since in that study

CDR interventions were found to be effective only when CDR committees were explicitly tapped to solve problems. But again there are no results in the existing literature on relative benefits of including or seeking to bypass existing powerbrokers.

Overall, even without considering normative arguments in favor or against interventions of the form favored by development actors, our findings suggest a need to rethink the idea of targeting local institutions for change. Our results suggest that the conceptualization of how local societies function are likely flawed, as are beliefs regarding the effects that efforts to change them might have. While substantive and intrinsic arguments for bringing aid allocation decisions to local levels may still justify the use of this type of intervention, this research provides grounds to drop the transformative ambitions.

References

Acemoglu, D., Johnson, S., and Robinson, J. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(5):1369–1401.

Acemoglu, D., Reed, T., and Robinson, J. A. (2014). Chiefs: Elite Control of Civil Society and Economic Development in Sierra Leone. *Journal of Political Economy*, (Forthcoming).

Aronow, P. M. and Samii, C. D. (2013). Estimating Average Causal Effects Under General Interference. *Working paper.*

Autesserre, S. (2010). *The Trouble with the Congo. Local Violence and the Failure of International Peacebuilding.* Cambridge University Press, New York City.

Avdeenko, A. and Gilligan, M. J. (2014). International Interventions to Build Social Capital: Evidence From a Field Experiment in Sudan. *World Bank Policy Research Working Paper Series*, 6772.

Barrios, T., Diamond, R., Imbens, G. W., and Kolesár, M. (2012). Clustering, Spatial Correlations and Randomization Inference. *Journal of the American Statistical Association*, 107(498):578–591.

Beath, A., Christia, F., and Enikolopov, R. (2011). Winning Hearts and Minds through Development Aid: Evidence from a Field Experiment in Afghanistan. *Working paper.*

Beath, A., Christia, F., and Enikolopov, R. (2013). Do Elected Councils Improve Governance? Experimental Evidence on Local Institutions in Afghanistan. *Working paper.*

Bidner, C. and Francois, P. (2013). The Emergence of Political Accountability. *Quarterly Journal of Economics*, 128(3):1397–1448.

Binmore, K. G. (1998). *Game theory and the social contract: just playing*, volume 2. Mit Press.

Bowers, J. and Fredrickson, M. M. (2013). Reasoning about Interference Between Units: A General Framework. *Political Analysis*, 21:97–124.

Bowles, S. and Gintis, H. (2004). The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations. *Theoretical Population Biology*, 65(1):17–28.

Casey, K., Glennerster, R., and Miguel, E. (2013). Reshaping Institutions: Evidence on Aid Impacts using a Preanalysis Plan. *Quarterly Journal of Economics*, 127(4):1755–1812.

Chen, J., Humphreys, M., and Modi, V. (2010). Technology Diffusion and Social Networks: Evidence from a Field Experiment in Uganda. *Working paper*.

Chwe, M. S.-Y. (2000). Communication and Coordination in Social Networks. *Review of Economic Studies*, 67(1):1–16.

De Angelis, C. D., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Marusic, A., Overbeke, A. J. P., Schroeder, T. V., Sox, H. C., and Weyden, M. B. V. D. (2005). Is This Clinical Trial Fully Registered? A Statement From the International Committee of Medical Journal Editors. *Canadian Medical Association Journal*, 172(13):1700–1702.

Fearon, J. D., Humphreys, M., and Weinstein, J. M. (2009). Can Development Aid Contribute to Social Cohesion after Civil War? Evidence from a Field Experiment in Post-Conflict Liberia. *American Economic Review: Papers & Proceedings*, 99(2):287–291.

Fisher, R. A. (1935). *The Design of Experiments*. Oliver and Boyd, London.

Freire, P. (1970). *Pedagoy of the Oppressed*. Herder and Herder, New York City.

Gandhi, M. K. (1962). *Village Swaraj*. Princeton University Press, Princeton.

Gerber, A. S. and Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton, New York City.

Hirschman, A. O. (1984). *Getting Ahead Collectively: Grassroots Experiences in Latin America*. New York: Pergamon Press.

Ho, D. E. and Imai, K. (2006). Randomization Inference With Natural Experiments: An Analysis of Ballot Effects in the 2003 California Recall Elections. *Journal of the American Statistical Association*, 101(475):888–900.

Humphreys, M., Sanchez de la Sierra, R., and Van der Windt, P. (2013). Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration. *Political Analysis*, 21(1):1–20.

IRC (2007). *Mortality in the Democratic Republic of Congo: An Ongoing Crisis*.

IRC (2012). Tuungane Phase I (2007-2010): IRC's Achievements in Community Driven Reconstruction in DR Congo. Technical report.

Kaplan, R. D. (1994). The Coming Anarchy: How Scarcity, Crime, Overpopulation, Tribalism, and Disease are Rapidly Destroying the Social Fabric of Our Planet. *Atlantic Monthly*, (February).

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, 75(1):83–119.

La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2008). The Economic Consequences of Legal Origins. *Journal of Economic Perspectives*, 46(2):233–285.

Logan, C. (2013). The Roots of Resilience: Exploring Popular Support for African Traditional Authorities. *African Affairs*, 112(448):353–376.

Mamdani, M. (1996). *Citizen and Subject: Contemporary Africa and the Legacy of Late Colonialism*. Princeton University Press, Princeton.

Mansuri, G. and Rao, V. (2004). Community-Based and -Driven Development: A Critical Review. *The World Bank Research Observer*, 19(1):1–39.

Mansuri, G. and Rao, V. (2013). *Localizing Development: Does Participation Work?* World Bank Policy Report.

Miguel, E. and Kremer, M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, 72(1):159–217.

Murphy, W. P. (1990). Creating the Appearance of Consensus in Mende Political Discourse. *American Anthropologist*, 92(1):24–41.

Nest, M. (2011). *Coltan*. Polity, New York City.

North, D. C. (1991). Economic Performance Through Time. *The American Economic Review*, 84(3):359–368.

Nunn, N. (2008). The Long Term Effects of Africa's Slave Trades. *Quarterly Journal of Economics*, 123(1):139–176.

Prunier, G. (2009). *Africa's World War: Congo, the Rwandan Genocide, and the Making of a Continental Catastrophe*. Oxford University Press, Oxford.

Putnam, R. D. (1993). *Making Democracy Work. Civic Traditions in Modern Italy*. Princeton: Princeton University Press.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5):688–701.

Schatzberg, M. G. (1997). Beyond Mobutu: Kabila and the Congo. *Journal of Democracy*, 8(4):70–84.

Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition have Failed*. Yale University Press, New Haven.

Sen, A. (2001). *Development as Freedom*. Oxford University Press, New York City.

Sinclair, B., McConnell, M., and Green, D. P. (2012). Detecting Spillover Effects: Design and Analysis of Multi-level Experiments. *American Journal of Political Science*, 56(4):1055–1069.

Small, D. S., Ten Have, T. R., and Rosenbaum, P. R. (2008). Randomization Inference in a Group-Randomized Trial of Treatments for Depression. *Journal of the American Statistical Association*, 103(481):271–279.

Sokoloff, K. L. and Engerman, S. L. (2000). History Lessons Institutions, Factor Endowments, and Paths of Development in the New World. *Journal of Economic Perspectives*, 14(3):217–232.

Turley, T., Voors, M., Bulte, E., Kontoleon, A., and List, J. A. (2014). Chief for a Day! Chiefs, Participatory Development and Elite Capture in Sierra Leone. *Working paper*.

Vlassenroot, K. and Huggins, C. (2005). Land, Migration and Conflict in Eastern DRC. *From the Ground Up: Land Rights, Conflict and Peace in Sub-Saharan Africa*, pages 115–194.

White, H. (1999). *Politicizing Development? The Role of Participation in the Activities of Aid Agencies*. Kluwer Academic Press, Boston.

World Bank (2009). IDA at Work Community-Driven Development: Delivering the Results People Need. Technical report.

Young, H. P. (2001). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, Princeton.

Supplementary Material

A Original Hypotheses

Table 3 lists the core hypotheses and the date of their generation. In the main text we discuss outcomes in terms of context and not in the order of the hypotheses provided here; however Table 2— provides a mapping between each of our tests and the corresponding hypothesis.

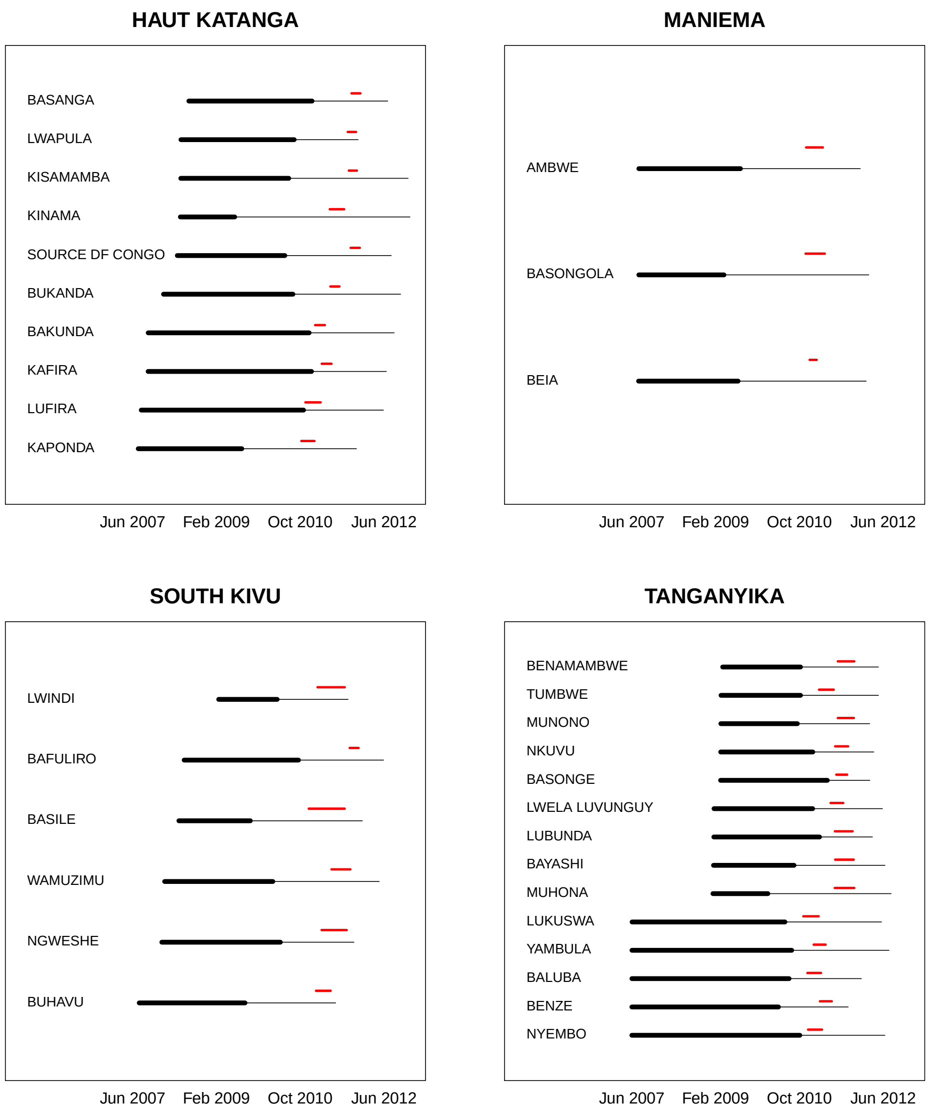
Table 3: Hypotheses Developed Prior to Data Collection

#	Category	Hypothesis	Date
H1	Accountability	Communities will be more proactive in seeking support from local government and NGOs for community initiatives and the private sector.	2007
H2	Participation	Individuals in *Tuungane* communities will report a greater sense of a right to take part in local decisions.	2007
H3	Participation	Individuals in *Tuungane* communities will report a greater sense of obligation to take part in local decisions.	2007
H4	Transparency	Individuals in *Tuungane* communities will report greater knowledge about local decision-making processes and outcomes.	2007
H5	Accountability	Individuals in *Tuungane* communities will report an increased willingness to hold traditional and political leaders accountable.	2007
H6	Participation	Individuals in *Tuungane* communities are more likely to believe that local leaders should be elected rather than selected through an alternative mechanism.	2007
H7	Efficiency	Projects will be implemented more efficiently in *Tuungane* areas.	2010
H8	Capture	Benefits will be more broadly distributed in *Tuungane* communities	2010

Notes: Hypotheses H1-H6 were generated in 2007 prior to the intervention. These hypotheses were then refined in 2010, before data collection. The refinements are found in (Ref Omitted).H7 and H8 were added in that phase. Other hypotheses related to intended effects on economic outcomes and social cohesion, as well as unintended consequences of various forms, are also in the design document and were registered at (Ref Omitted) prior to data collection.

B Timing of Intervention and Measurement

Figure 3: Timeline of Implementation



Notes Thin black lines indicate length of the *Tuungane* project per chiefdom. Thick line indicates the first (VDC) phase. Shorter, red lines indicate the period of measurement in that chiefdom. Source: Authors' drawing.

C Balance

The analyses in this paper relies on randomization which guarantees that treatment and control units will be similar in expectation. In practice, however, it is possible for treatment and control units to differ simply by virtue of unlucky draws. To test this we analyze the following variables at the village level: distance from major urban center, village population size, prior poverty, exposure to conflict, existence of prior NGO activity; and the following pre-treatment individual level variables: gender, age, education, migration status. These variables were pre-specified and are not selected based on the strength of their correlation with treatment (see (Ref Omitted)). Table 4 lists the average for each variable for the *Tuungane* and the control group, and the difference between both. We find that there are no strong differences across these two groups, which is consistent with what is to be expected given the random assignment.

Table 4: Balance Test

Variable	Level	*Tuungane*	Control	Difference?	N
Distance from major urban center	Village	8.99	8.99	0.01	804
Population size of village	Village	488.35	469.98	18.37	457
Prior level of poverty	Village	3.18	3.42	-0.24	710
Exposure to conflict	Village	2.43	2.44	-0.01	992
Existence of prior NGO activity	Village	0.43	0.42	0.01	992
Gender ratio	Individual	0.51	0.50	0.01	5,539
Age	Individual	39.26	39.73	-0.47	5,409
Education	Individual	4.35	4.35	-0.01	3,978
Migration status	Individual	0.46	0.47	-0.01	3,733

Notes: Based upon the following measures: QE13E, AC11, QC23-27, CQ39 (2007 baseline survey), CQ68 (2007 baseline survey), QF7, QF9, QF13, SP1. Exposure to conflict and existence of prior NGO activity have been aggregated to the chiefdom level. Comparing treatment and control communities taking into account weights and clustering gives the same result.

As described in our pre-analysis plan (Ref Omitted) we can also control for these key variables that are plausible related to outcomes even though we do not expect them to be related to treatment. In so doing we can reduce variance and generate more precise estimates of effects as well as correct for random imbalances. Redoing the analysis in this paper controlling for the variables listed in Table 4 does not change our results (not reported).

D RAPID Script

We provide below the full text of the description of RAPID to communities during the general Assembly meeting in Step A of the RAPID process.

"I work for RAPID and I want to talk with you about a project that we are introducing in this village. RAPID, which stands for "Research-Action through Projects for Development Impacts." The project provides development funding from the British government and is coordinated with researchers from Columbia University in New York and from the universities of Bukavu and Lubumbashi. The aim of the project is to provide development aid to your community while at the same time contributing to scientific research to better understand your priorities and needs.

Your village and other villages were selected in a lottery involving all the villages in this territory for the program. The program will provide a grant of at least $900 (perhaps more) in international funding to implement a quick impact project. In this project we will let the community decide how best to use the funds.

Your Chief [name] gave us permission to hold this meeting as a prerequisite for participation in the project. The aims of this meeting are to inform you of the program, to provide you the opportunity as members of the village to ask us any questions about the project, and to offer a forum for discussion on development priorities in this village and use of these funds.

There are a few requirements for participation in this project, and it is important to us that you understand them:

1. First, we want the community to decide how to use the project funds. Following this meeting, your village will have seven days to decide how to use the funds. The total funding guaranteed for this community is at least $900. It is up to you as a village to decide the best use of funds. There are no restrictions on the use of funds, except they must be used to benefit the community and be spent out by you in the next 50 days. For this reason we encourage you to use the funds to assist members of the community through projects such as purchasing and distributing seeds, tools, large participatory work or other

projects that support the well-being of this community. These funds may also be distributed to community members to use at their discretion. We prohibit the use of these funds to purchase any item whose purpose is to harm others.

2. Second, we are asking the community to identify people to represent the village for this project. These individuals will be responsible for carrying out the accounting of the use of these funds. It is up to the community to decide who these people will be over the next seven days. You are free to choose any person or persons that you feel are most appropriate to act as representatives.

3. Third, we ask you to complete this form [show the form] to return it on [date]. It is the Project Description Form. [Show form BP1]. I will leave it with you today to complete over the next six days. The information in the form will contain the decisions you have made for the project. A representative of Project RAPID will return in six days to collect this form. We will not be able to make the grant payment if you do not complete this form.

4. Fourth, among the questions I ask you to fill out on the form are: who are the individuals who will be responsible for managing these funds?; which project the community has chosen?; and what is the budget of such a provisional project?

5. Fifth, we ask that in two months, representatives of the community for the project RAPID provide us with an accurate accounting of the usage of funds, with evidence. This is to facilitate our understanding of the priorities of your village, as part of our research.

6. Finally, in accepting this project you also accept that the use of Project RAPID funds will be subject to an audit. What will this look like? We will send teams to implement an audit in certain villages participating in the program: if this village is audited, we will examine what the village has done with project funds. The findings will contribute to our study of the needs of Eastern Congo.

Information on the disbursement of funds will be provided when collecting Project Description Forms from the representatives chosen by the community for the management of funds. Following receipt of these funds, your village will spend out these funds for your chosen project over the next 7 weeks (49 days), as is compatible with the project.

Do you have questions about this process? Would you like to participate in this project?

As we said before, there is a research component linked with this project. It is important for us that you have a good understanding of what is involved in this research so that you can use that understanding either agree or refuse to take part in it. As this project is implemented we will seek to hold a series of interviews with members of this community. These interviews will all be anonymous interviews. The aim of these is to understand the communitys priorities. It is important that you understand that if you choose to be interviewed your responses will be kept anonymous.

Another part of our research will be on decision making during community meetings. Collecting measures during discussions helps us to understand more about this community and its priorities. Again we will only do this if the community agrees to this and in all cases information that is recorded will be done in a way that conserves anonymity.

Before asking for your consent we want to note that this research does not bring risks, but nor does it bring direct benefits for you. By improving our understanding of community priorities in East Congo this research seeks to contribute to an improvement in the quality of development aid throughout the area.

Do you consent to us collecting this data to help with this research?"

E Robustness

Table 5: Results Robustness Tests

Concept	Measure (#)	Base	Alt. Treat.	Spillover Effects D(5km)	I(5km)	D(20km)	I(20km)	Alt. Specifications Village	Bins	Prop.
Participation	Meeting Att. (1)	-1.98	-3.98	-5.39	22.10	0.09	-2.80	-1.98	-1.24	-1.56
		(0.79)	(0.59)	(0.809)	(0.517)	(0.495)	(0.964)	(0.79)	(0.83)	(0.84)
	Interventions (1)	-0.49	-0.19	-0.42	-0.07	-0.41	-0.29	-0.49	-0.37	-0.42
		(0.25)	(0.66)	(0.722)	(0.262)	(0.815)	(0.387)	(0.25)	(0.31)	(0.32)
	Dominance (1)	0.52	-0.33	0.52	-1.27	0.04	0.14	0.52	0.52	0.74
		(0.73)	(0.82)	(0.374)	(0.356)	(0.475)	(0.195)	(0.73)	(0.68)	(0.62)
	Selection Meth. (2)	0.07	0.05	0.12	0.28	-0.03	0.41	0.07	0.07	0.08
		(0.45)	(0.59)	(0.232)	(0.308)	(0.62)	(0.289)	(0.45)	(0.30)	(0.40)
	Committee Comp. (3)	0.08	0.10	0.18	0.05	0.13*	0.24	0.08	0.09	0.07
		(0.42)	0.1	(0.056)	(0.866)	(0.072)	(0.134)	(0.42)	(0.19)	(0.44)
Accountability	Accountability Mech. (4)	0.00	-0.02	0.01	0.09	-0.08	-0.17	0.00	0.00	-0.02
		(0.97)	(0.83)	(0.482)	(0.414)	(0.749)	(0.871)	(0.97)	(1.00)	(0.85)
	Complaints (5)	0.02	-0.01	0.13	0.24	-0.01	0.32	0.01	0.01	0.00
		(0.75)	(0.89)	(0.143)	(0.631)	(0.588)	(0.525)	(0.90)	(0.89)	(0.99)
	Complaints Cond. (5)	0.68**	0.71*	0.59*	0.25	0.59*	0.25	0.82**	0.57*	0.77**
		(0.07)	(0.06)	(0.09)	(0.54)	(0.09)	(0.54)	(0.02)	(0.05)	(0.03)
Efficiency	Accounting (6)	0.01	-0.05	-0.11	-0.36	0.00	-0.61	0.01	0.01	0.01
		(0.90)	(0.63)	(0.678)	(0.576)	(0.493)	(0.581)	(0.90)	(0.88)	(0.95)
	Info Transm. (7)	-1.41	0.08	-2.37	-1.14	-2.37	-1.14	-0.61	-0.74	-0.98
		(0.37)	(0.96)	(0.23)	(0.55)	(0.23)	(0.55)	(0.67)	(0.61)	(0.50)
Transparency	Knowledge (8)	1.52	-0.73	-2.98	-9.72	0.26	-12.16	1.27	1.44	1.69
		(0.64)	(0.80)	(0.697)	(0.3)	(0.467)	(0.412)	(0.66)	(0.52)	(0.55)
	Seek Info (9)	3.84	4.01	5.86	2.13	0.43	12.47	2.19	1.67	1.35
		(0.24)	(0.16)	(0.109)	(0.789)	(0.419)	(0.12)	(0.48)	(0.45)	(0.66)
Capture	Fin. Irregularities (10)	3.52	-13.94	-29.16	-13.57	-0.25	-46.86	3.52	5.53	1.19
		(0.87)	(0.50)	(0.847)	(0.248)	(0.507)	(0.456)	(0.87)	(0.76)	(0.96)
	Beneficiaries (11)	3.28	6.41	4.66	-3.13	1.12	11.99	3.09	-1.98	3.05
		(0.55)	(0.21)	(0.344)	(0.902)	(NA)	(NA)	(0.54)	(0.63)	(0.54)
	Inequality (11)	0.56	-0.14	-0.72	2.02	-0.77	1.68	-0.15	0.58	-0.11
		(0.71)	(0.92)	(0.694)	(0.23)	(NA)	(NA)	(0.92)	(0.65)	(0.94)
	Dominance of Pref. (12)	-0.01	0.00	-0.03	-0.01	-0.03	-0.01	-0.01	-0.01	-0.02
		(0.73)	(0.92)	(0.30)	(0.84)	(0.30)	(0.84)	(0.66)	(0.64)	(0.58)

Notes: One, two or three asterisks indicate, respectively, significance levels at the 10%, 5% and 1% at one-sided tests. P-values are reported in parentheses. 'Base' corresponds to the results reported in Table 2. "Alt. Treat." are results using a treatment variable that uses IRC's classification of treatment in cases in which databases disagreed. D() indicates the weighted average of the direct *Tuungane* effect for those village indirectly and not indirectly treated. I() indicates the the weighted average of the indirect *Tuungane* effect for those village that are directly treated and those that are not. Results on direct and indirect effects for rows that involve an interaction effect correspond to β_4 and β_5 of the following regression $Y = \alpha + \beta_1 D + \beta_2 I + \beta_3 X + \beta_4 D * X + \beta_5 I * X$, where D (I) indicates whether the unit was directly (indirectly) treated, and X is the conditioning variable. We cluster the errors at the CDC level and use weights that take into account the probability of each unit being directly/indirectly treated. "NA" is reported in brackets for Beneficiaries and Inequality because of the low number of observations. "Village" are results in which all variables are aggregated to the village level using individual sampling weights. "Bin" are results at the village level introducing controlling for lottery bins. "Prop." are results (at the village level) using propensity weights adjusted to assess VDC level sample average treatment effects rather than population average treatment effects. In order to aggregate the dominance of preferences measure to the village level, we construct a dependent variable by subtracting the average number of times individuals' preferences correspond to the project implemented (a measure between 0 and 1) from whether the chief's preference corresponds to the project (either 0 or 1).

F Social Desirability Bias

Table 6: Social Desirability Test

	Negative Prompt	**Positive Prompt**	**Difference (se)**
Control	0.641	0.833	0.192***
(n)	955	959	0.025
Tuungane	0.634	0.850	0.216***
(n)	939	952	0.027
Difference	-0.008	0.017	0.024
(se)	0.030	0.022	0.037

Notes: Share of individuals answering ‘yes’ to the question “Do you agree with the view that elections are the best way to choose community representatives to serve in positions that require technical expertise?” $***$ indicates significance at the 1% level. Results take into account sampling weights.