SOCIAL ENGINEERING IN THE TROPICS: A GRASSROOTS DEMOCRATIZATION EXPERIMENT IN THE CONGO

Macartan Humphreys* Raul Sanchez de la Sierra† Peter Van der Windt‡

Version 2.1: May 2015

Abstract

Recent scholarship has provided mixed evidence on the ability of international interventions to foster local democratic practices in post conflict environments. We bring new evidence to the question using an experiment that replicates and extends past attempts to answer this question. Studying an intervention that is larger in scale than those previously examined, we employ a new set of behavioral measures that more closely track outcomes of interest, and exploit design based estimation techniques that seek to assess bias due to researcher demand effects as well as treatment spillovers. Across five families of outcome measures we find almost no evidence to support the idea that external interventions have a substantial impact on local governance structures in the ways currently posited by international actors.

JEL Codes: D72; P48; D02; O17

Keywords: Political Processes; Political Economy; Institutions: design, formation, operations, and impact; Demonstration Effects

*Columbia University. Corresponding author: mh2245@columbia.edu. This research was funded by the International Initiative for Impact Evaluation (3IE) and the Department For International Development, UK. We thank the International Rescue Committee and CARE International for their partnership in that research. Humphreys thanks the Trudeau Foundation for support while this work was undertaken.

†Harvard University. rsanchezdelasierra@fas.harvard.edu.

‡New York University - Abu Dhabi. petervanderwindt@nyu.edu

1 Introduction

A large body of research suggests that institutions are a key driver of economic development (Sokoloff and Engerman, 2000; Acemoglu, Johnson and Robinson, 2001; La Porta, Lopez-de Silanes and Shleifer, 2008). According to North (1991), for example, "institutions form the incentive structure of a society and the political and economic institutions, in consequence, are the underlying determinant of economic performance." In recent years international aid has taken a cue from this proposition and sought to jump-start development processes by introducing institutional innovations in hopes that these will lead to greater local accountability and persistent social change.[1] The rapid growth of this approach reflects the widespread belief that institutions matter at least as much as fundamentals and that exposure to inclusive institutional practices can create demonstration effects that in turn lead to institutional uptake and persistence without any changes to fundamentals.

However despite the popularity of the underlying theory and the large amounts of development aid allocated to interventions of this form, past studies have found only mixed evidence that these intervention produce the benefits claimed of them. As we review below, some studies have found evidence of expected effects in specific areas or for specific subgroups (Fearon, Humphreys and Weinstein, 2009) but many have failed to find effects on key outcomes and on political accountability in particular (Casey, Glennerster and Miguel, 2013; Avdeenko and Gilligan, N.d.; Beath, Christia and Enikolopov, 2013; King and Samii, 2014). If indeed these interventions make little or no dint on actual behaviors, this has implications for development theory and practice. For development theory it supports the idea that institutional arrangements reflect more fundamental relations of power and are not an important independent source of growth; for development practice it supports views that development interventions should focus on economic fundamentals rather instead of trying to alter political behavior directly.

We contribute to this literature through replication and extension. Convinced of the importance of knowledge accumulation in this area, we examine a grassroots democratization intervention that tracks closely in purpose and design the interventions studied in past work. We see the replication component of this research as an important part of our contribution. Like past studies, we rely on random assignment to treatment to achieve causal identification; in this case with 1,250 villages of eastern Democratic Republic of Congo (DRC) randomly selected to participate in a four year long "Community Driven Development" intervention (CDD). During this intervention, populations participated in elections, made decisions about

[1]Since the 1990s, "participatory development" has become a favored model for development, and has formed a major pillar of post-conflict development Mansuri and Rao (2013) quote a figure of $85bn in World Bank spending in the last decade alone on this broad class of interventions.

resource allocations democratically, and used practices designed to hold leaders to account for mismanagement. With many past results tending negative, our study examines a site, Eastern Congo, that is a particularly well-suited region to expect positive effects. In this sense it is a hard case for the null proposition. The institutional context of East Congo is generally described as plagued by capture by traditional elites, predatory state rule, pervasive corruption, and undemocratic. This context is representative of a large number of post-conflict countries (Acemoglu, Reed and Robinson, 2014) with great scope for institutional gains.

Beyond replication, our approach advances the existing literature on five dimensions. First, we measure institutional change by collecting information on behavior and distribution of economic benefits in a naturalistic context via observation of community behavior in a real collective allocation problem. Close to the strategy used in Beath, Christia and Enikolopov (2013), this stands in contrast to the survey approach used in Barron et al. (2009), the behavioral games used in Fearon, Humphreys and Weinstein (2009) and Avdeenko and Gilligan (N.d.) or the framed exercises used in Casey, Glennerster and Miguel (2013). Drawing on a mix of audit data, direct observation, household surveys before and after the cash transfer, we generate behavioral measures for five core dimensions of democratic governance that this intervention sought to affect: *participation*, *accountability*, *efficiency*, *transparency*, and *capture*.

Second, we employ data from over 800 villages in 456 clusters, an unusually large sample that significantly reduces the likelihood of false negatives. This compares to 83 villages in Fearon, Humphreys and Weinstein (2009) and 236 villages in Casey, Glennerster and Miguel (2013) and makes this intervention one of the largest field experiments of its kind.

Third, we provide one of the first applications of a design-based strategy for assessing spillover effects in this field. Spillover effects present a serious challenge to assessments of causal effects and may be a particular concern in our case.

Fourth, we disclosed our analysis plan prior to data collection. Our analysis joins a still small number of studies that have publicly posted and subsequently sought to follow a detailed pre-analysis plan. In contrast to these studies, however, we specified not just our core hypotheses, but also the details of all major analyses and tests we run in this paper. This protects us from data mining, not only to select narratives, but also to select measures, subsets of the data, or interactions that may generate results.[2]

Finally we include in our analysis a set of innovations in measurement designed to better understand the null affect with the aim of assessing not simply whether there is evidence

[2] See Casey, Glennerster and Miguel (2013) for further discussion of pre-analysis plans. See Online Appendix K for a description of deviations we made from our plan and the implications of these deviations.

for effects but if not why not. Unusually, one of these is a measurement of the *priors*: before generating data on outcomes we gathered information on the beliefs of researchers and development actors regarding the likely effect of this program; this data on priors allows for a clearer assessment of the contribution of our findings to knowledge on the effects of aid.

Our results consolidate a view emerging from prior studies: across a wide array of measures we find little or no evidence that exposure to democratic practice has any effect on power structures, either in treated villages or in neighboring villages. Our preferred explanation for this is the simple one that these effects are not in operation and that the optimism of international organizations is misplaced. We support this view by examining whether the failure to measure effects could be due to poor project implementation, the operation of positive spillovers, poor measurement, heterogeneous effects, elite backlash against democratization, or changes in expectations about future aid in control villages.

2 A Grassroots Democratization Experiment

2.1 Theory and Evidence

The idea that efforts to import institutional innovations can produce large social change is consistent with many accounts of historical patterns of economic development as well as a large body of theory on social change.

In the account of Rodrik, Subramanian and Trebbi (2004), "institutions rule" and variations in institutions trump all other prominent explanations of macro-level (cross national) variation in economic success (see also Weingast (1995), La Porta, Lopez-de Silanes and Shleifer (2008) and Acemoglu, Johnson and Robinson (2001)).In these accounts, institutions alter behavior by altering expectations (or in a sense made more precise below, institutions *are* expectations). Thus even accidental changes in institutions, that are not accompanied by other changes, can have large effects. This feature of these arguments is in evidence in the identification strategies employed. In Acemoglu, Johnson and Robinson (2001) for example the use of settler mortality as an instrument presupposes that any structural changes associated with colonial presence (such as access to technologies, access to trade, or changes in demography or human capital) do not have direct effects. Similarly in Rodrik et al's account, while geography may account for institutional variation it does not have an independent causal effect (Rodrik, Subramanian and Trebbi, 2004). For a general discussion and criticisms of this macro level research see Glaeser et al. (2004).

These accounts are consistent with theoretical work that adopts a conceptualization of institutions as an equilibrium of a larger game (Shepsle, 2006; Greif and Laitin, 2004) and that highlights how institutional variation may reflect differences in equilibrium selection, rather than differences in primitives (i.e. variation arises from the choice of equilibrium rather than the sets of equilibria available). Young (2001) for example provides an account of social institutions as patterns of behavior that may exhibit large variations across space and time without any change to fundamentals. In some accounts, seemingly deep social structures coupled with policy trends (such as patterns of class identification coupled with redistributive policies) can obtain as equilibriums in environments where very different equilibriums also obtain, supported by the same fundamentals (such as weak class identification and low levels of redistribution) (Shayo, 2009). In other accounts, variation in the quality of property rights regimes, norms of fairness, or tolerance for more or less accountable governments is a function of equilibrium selection rather than fundamentals (Grossman and Kim, 1995; **?**). In some environments, power structures are themselves a property of equilibrium: if actors coordinate strategies in one way, the decisions of traditional institutions will be treated as authoritative, but in other available equilibriums, they would be subject to challenge (see for example Young (2001); Chwe (2000); Bidner and Francois (2013)).

The distinction between changes in outcomes due to changes in institutions proper and changes due to changes in the way institutions affect outcomes when fundamentals change can be a subtle but important one. In Supplementary Materials A we use a simple game to describe two intervention strategies that lead to observationally equivalent behavior but through distinct mechanisms. One, "Strategy A", affects behavior by altering expectations directly; the other, "Strategy B", alters behavior by altering fundamentals.

Pure institutional interventions, often operating at the micro level, seek to change outcomes using an approach similar to Strategy A. Such approaches are from one perspective very optimistic. They make use of the fact that although patterns of weak accountability and poor growth are robust in the sense of arising from non-cooperative equilibriums (no one has unilateral incentives to change behavior) they are fragile in the sense that dramatic behavioral shifts can arise due to changes in expectations only. Although in the discussion above of Strategy A there was little formal specification for *why* external interventions would be successful at generating these shifts in equilibrium, lab experimental evidence demonstrates clearly that in some contexts external actors can use equilibrium-irrelevant interventions (such as labeling options or framing the context) to alter the focality of different equilibriums (Mehta, Starmer and Sugden, 1992). Bidner and Francois (2013) provide a more developed approach in the context of a model of accountability relations in which changes in norms occur endogenously following particular sequences of actions by leaders. Such results suggest that short run interventions that alter the behavior of leaders could lead to long run changes in equilibrium play. Similar results would obtain from the limited rationality models in Young (2001) where expectations are based on observation of past actions and equilibriums could change following a period of induced deviations.

Note that the focus on multiple equilibriums of Strategy A resembles "poverty trap" arguments that are used to support a "big push" approach to development. These accounts suggest that an externally produced change in equilibrium can produce sustainable change (Sachs, 2005). There is a critical difference however: in poverty trap arguments, equilibriums are differentiated not just by actions and expectations, but by *factors*, such as the stocks of human or physical capital available to an economy. In practice then these poverty trap arguments call for strategies more like Strategy B than Strategy A.

This optimistic account underpinning Strategy A faces many challenges however. The first is that even if purely institutional accounts such as those in Acemoglu, Johnson and Robinson (2001) are historically correct; the fact that changes in outcomes are due to changes in expectations only does not mean that these changes are *easy* to produce. Most of the theory says little about whether shifts between equilibriums are easy to produce. The second is that in many accounts of changes that seem to support a logic for interventions following Strategy A, processes more like those assumed by Strategy B may be in operation. Thus many historical accounts that emphasize the importance of institutions, give primacy to more fundamental features. Sokoloff and Engerman (2000) in their study of the role of institutions in maintaining inequality emphasize the importance of factor endowments in determining structural inequality; indeed they highlight the "clear implication that institutions should not be presumed to be exogenous." In the account provided by Herbst (2014), institutional variations in state capacity and responsiveness also reflect more fundamental features, notably agricultural technology and population densities. Other accounts emphasize access to resources, such as subsoil resources (Ross, 2001) or aid (Nunn and Qian, 2014).

However, despite these challenges, the simple institutional account also resonates with views among policymakers about the promises of exposure to democratic practice. Mansuri and Rao (2013) discuss the motivations and learning from $85bn spent by the World Bank alone on local participatory government which has, as one of its primary aims "strengthening demand for good governance." Perhaps the most prominent model in this arsenal of interventions — and one that recent scholarship in economics and political science has focused on — is the use of 'Community Driven Development" (CDD) interventions to bring about local change. 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).

As described by King and Samii (2014) and others there is no single theory of *how* CDD is meant to produce a change in social behaviors, though many of the mechanisms described by practitioners focus on changes in expectations and practices arising from exposure to democratic institutions. These include introducing awareness of alternative strategies to decide on allocations, inducing expressions of aspirations among excluded populations, providing a focal point for the population to solve collective action problems. CDD interventions include an investment component though these play an instrumental role, providing an incentive for communities to engage in CDD programs. Indeed if in early models of CDD the investments constituted a large share of overall project budgets, a more recent focus on institutional concerns has occurred alongside a shift in budgets from projects to social activities.

Despite the popularity of the CDD model there has been little evidence for the claims made on its behalf (Mansuri and Rao, 2013). Recently, a number of studies have examined the social and economic effects of similar programs, but the picture they paint is inconclusive. Fearon, Humphreys and Weinstein (2009), examining a CDD intervention in Liberia, find mixed evidence of impacts on short term governance outcomes and positive evidence for impacts on social cohesion — as measured through the use of a public goods game. The evidence on effects for social cohesion is found for one of two measures only however, and, as the authors note, arises for a measure in which the use of the CDD institutions are encouraged. Beath, Christia and Enikolopov (2013) also find some mixed evidence from a study of a large intervention in Afghanistan. There they found positive effects on governance outcomes when use of CDD institutions was mandated in the behavioral measurement strategy, but no effects when these institutions were not mandated. An interpretation consistent with both sets of findings is that CDD created an institutional mechanisms, but not one that was in fact adopted by communities. Other studies found still weaker results. Casey, Glennerster and Miguel (2013) find almost no evidence for social change across a wide array of framed behavioral measures in Sierra Leone and Avdeenko and Gilligan (N.d.) found no evidence for social changes using experimental games in Sudan. Surveying other studies in Senegal, Indonesia and elsewhere, Wong (2012) describe little or no evidence for social impacts and mixed evidence on governance outcomes.

2.2 The Site

Our area of study — South Kivu, Maniema, Tanganyika and Haut Katanga — figured centrally in the violence that has engulfed the country over the last two decades. Located in southern and eastern Congo, 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 (IRC, 2007). Despite the formal end to the war in July 2003, much of the program area continues to experience conflict.

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 and reconstruction programs.

2.3 The Intervention

For our analysis we focus on a large CDD program, "Tuungane," that was implemented in 1,250 war-affected villages throughout Eastern Congo. With an average of around 1,424 inhabitants per village the program sought to reach a beneficiary population of approximately 1,780,000 people.[3] The program was implemented in about four years, with the phase we study in this paper finished after around two years. See Figure 2 in the online appendix for an illustration of the timing of implementation across areas. As stated in the original program description, *Tuungane* sought 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."[4]

To reach these goals the implementing agency undertook a number of key activities. Populations were trained and mobilized before organizing elections in which so-called Village Development Committees ("VDCs") were formed. Next, VDC members, in consultation with the population, decided how to allocate an envelope of \$3,000 for a maximum of two projects. The proposed project(s) was then voted on by the whole village. A key component of *Tuungane* was that the VDC was 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. Activities took place with tight monitoring by the implementing partner. On average, four general assemblies were convened by the VDCs to justify the use of program funds to populations. Furthermore, VDC members participated in intense training on their roles and responsibilities, on leadership, on principles of good governance and gender issues. Leaders were also trained on financial management and accounting practices. Finally, communities were expected to contribute to their chosen VDC project with cash or in-kind support.

[3]The program's budget was £30m (USD \$46m), which includes the cost of the larger infrastructure projects that are not part of this study.

[4]In 2007, in collaboration with the implementing partner, the research team developed hypotheses that took account of these goals (Humphreys, Sanchez de la Sierra and Van der Windt, 2011).The primary hypotheses on governance outcomes are shown in Table 8 in the online appendix. 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 Humphreys, Sanchez de la Sierra and Van der Windt (2011).

3 Empirical Strategy

3.1 Random Assignment

Communities were assigned to receive *Tuungane* randomly, through public lotteries. All communities were first clustered and grouped geographically into 83 "lottery bins" from which clusters of villages were randomly drawn.[5] In total 600 village clusters entered the lottery, 280 were selected for treatment and the remaining 320 were in control (see Figure 1). Randomization by lottery bins achieves geographic balance across treatment and control within lottery bins. This reduces the variance of our estimates. In addition, *public* lotteries have a number of implications. First, they provide a form of informed consent on the part of communities, both those that benefit from the program and those that do not. Second, there is transparency over the selection process and this reduces concerns that one community was being unfairly favored over another. Third, public lotteries could lead to jealousy which could lead to bias in estimates if for instance, control communities may have started performing better or worse as a result of not being chosen. Our investigations of perceptions of the lotteries suggest that this is not likely a large concern however.[6]

3.2 Outcomes and Measurement

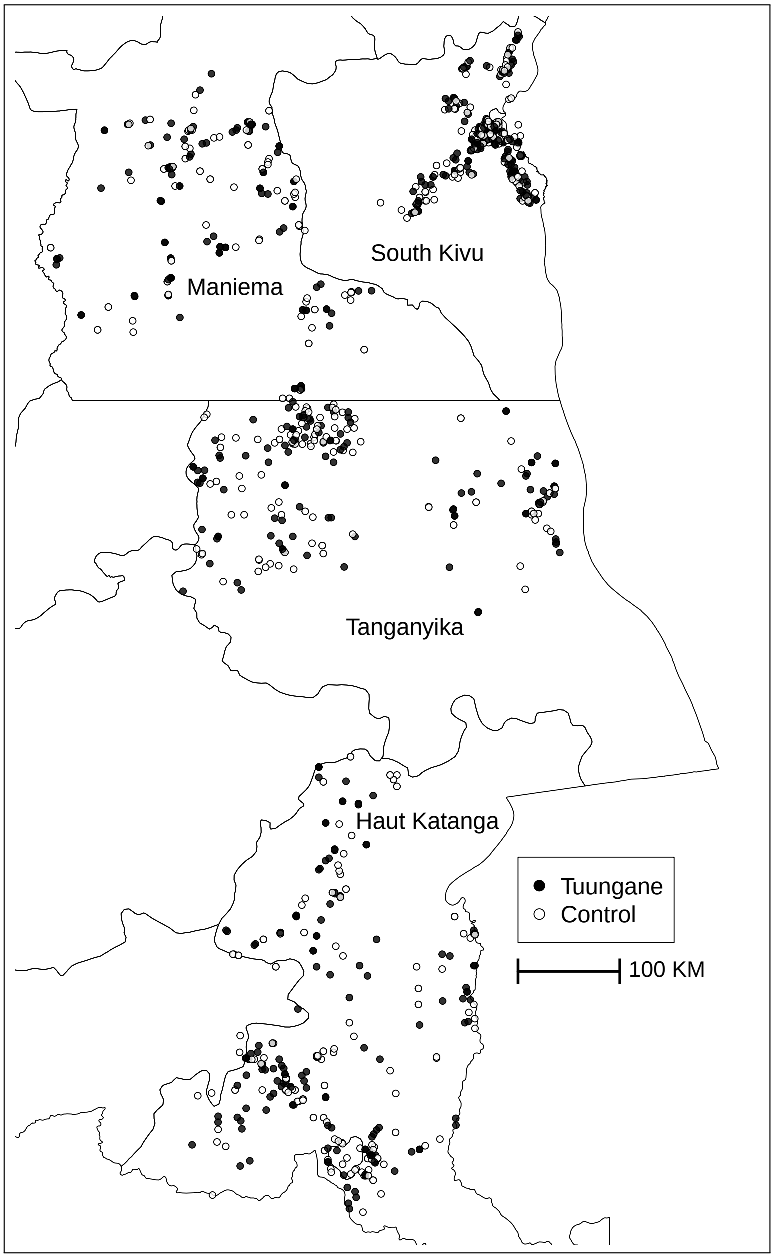
We sought to measure changes in governance behavior along five dimensions corresponding broadly to the effects of the intervention as described by the funders and implementers of the CDD program: *participation*, *accountability*, *efficiency*, *transparency*, and *capture*. We describe each of these in more detail below.

To measure the change in outcomes induced by demonstration effects, we confront measurement challenges. Survey data may be biased. For example, individuals in treatment (or control) communities may try to respond in ways that please outside funders. Lab-in-the-field type measures can suffer from an interpretation challenge — with these we might observe unbiased effects but those effects may be measured on a metric with no clear interpretation. To move beyond standard survey measures and lab-in-the-field style measures of behavior. We introduced an independent cash delivery project ("Recherche-Action sur les

[5] Lottery bins largely corresponded to chiefdoms ("Chefferies") or sectors ("Secteurs"). For simplicity, we generally use the term chiefdom for both units.

[6] 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 although in these areas the vast majority of respondents either had not heard of *Tuungane* or had no explanation for why the program was not implemented in their community.

Figure 1: Distribution of Treatment and Control clusters



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.

Projets d'Impact pour le Développement", henceforth RAPID) to assess behavioral change. 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 remaining 280 had not. Communities were unconstrained to identify and implement projects subject to minimal constraints.[7] While the RAPID project moderately encouraged distributive projects, these were not required. Importantly, the unconditional cash transfer left communities free to decide who should manage the funds and how decisions should be made. We rolled the RAPID project out in four steps (A-D) over the course of 2-3 months. The key features are described in Table 1; see the script provided in online appendix D.

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.

To measure these categories of behavior, we employ direct observations by enumerators of behavior in the village, extensive audits in each RAPID village, and large-n survey data collected during different steps of RAPID. We exploit sixteen core behavioral measures, organized into five groups. The first four of these groups — *participation*, *accountability*,

[7]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 — in order to be able to assess the use of funds in a timely manner.

efficiency, and *transparency* — are used to assess whether practices encouraged by the intervention were subsequently adopted. The last — *capture* — is used to assess overall effects on outcomes. We describe each measure in each family in turn in the next section.

3.3 Estimation

We compare mean outcomes in *Tuungane* communities with those in control communities, which, under conditions specified by (Rubin, 1974), provides an unbiased estimate of the average treatment effect. We account for small differences in assignment propensities in different randomization bins using inverse propensity weighting and employ sampling weights to account for differences is sampling probabilities reflecting differences in village sizes and differences in household sizes. Where individual level data is used estimates are clustered at the level of treatment (village clusters).

For some analyses we have access to multiple measures.[8] In order to generate a meaningful summary of multiple effects within each family, we follow the approach of Kling, Liebman and Katz (2007) and create standardized indices of outcomes on related items.

4 Main Results

Overall we find that exposure to grassroots democratization left power structures and related behaviors unaffected. Table 2 provides the results (see also Figure 3 in Supplementary Materials for a graphical representation of these results). Here the "Control" column describes the estimated level for each measure in control communities. Due to randomization, this is the expected outcome in the absence of the program. We provide the estimated effect of *Tuungane* in the subsequent column.[9]

4.1 Participation

A set of participation measures captures the extent to which communities adopted practices intended to provide formal scope for broad input to decision making. These include measures

[8] This raises concerns about interpretation. 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.

[9] Given the hypotheses of the program, these tests are conducted as "one-sided tests" — we thus provide the results of the test of whether there is sufficient evidence to reject the hypothesis that the program did not have any positive effect.

Table 2: Main Results

Concept	Measure	Index	Control	Effect	(se)	N
Participation	1 Meeting Attendance	No	130.48	-1.98	(7.40)	455
	2 Interventions in Meeting	No	14.21	-0.49	(0.42)	457
	3 Dominance in Discussion	No	70.81	0.52	(1.49)	457
	4 Participatory Selection Methods	Yes	0	0.07	(0.09)	451
	5 Committee Composition	Yes	0	0.08	(0.10)	452
Accountability	6 Accountability Mechanisms	Yes	0	0.00	(0.10)	414
	7 Private Complaints	Yes	0	0.02	(0.07)	3647
	8 Private Complaints | Management	Yes	0.29	0.68	(0.38)	3502
Efficiency	9 Quality of Accounting	Yes	0	0.01	(0.11)	399
	10 Information Transmission	No	9.66	-1.41	(1.56)	3800
Transparency	11 Knowledge of Project Amount	No	38.60	1.52	(3.21)	3685
	12 Willingness to Seek Information	No	37.70	3.84	(3.28)	1406
Capture	13 (Fewer) Financial Irregularities	No	851.51	3.52	(20.74)	394
	14 Number of beneficiaries	No	40.95	3.28	(5.52)	154
	15 Inequality of benefits	No	8.59	0.56	(1.52)	127
	16 Dominance of Preferences	No	0.04	-0.01	(0.03)	2666

Notes: For a more complete discussion on each measure, see Humphreys, Sanchez de la Sierra and Van der Windt (2012). For "Private Complaints | Management", "Information Transmission" and "Dominance of Preferences" we estimate $Y = \beta_0 + \beta_1 X + \beta_2 T + \beta_3 XT$, and report β_1 in the control column and β_3 in the treatment effect column where X corresponds, respectively, to fund mismanagement, "RAPID," and "chief." All analyses employ propensity score weights and clustering of standard errors at the level of randomization clusters.

of behavior in general assemblies, including how many took part, who spoke, and whose voice dominated; measures of voting mechanisms used to select committee members and projects; and measures of the size and gender composition of RAPID committees allows us to construct a measure of inclusion at the village level.

We begin with the examination of villagers' attendance and patterns of social interaction in the first village meeting. Table 2 ("*Meeting Attendance*") shows that in control communities on average 130 adults participated in the public meeting, two more than in *Tuungane* communities, a very small difference which is not statistically significant. The second row ("*Interventions in Meetings*") provides results for the patterns of social interaction. On average, fourteen interventions are made per meeting, with only marginally fewer interventions in *Tuungane* communities. Furthermore, we 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

("*Dominance in Discussion*") shows that the patterns of dominance of social interactions are indistinguishable across treatment and control.[10]

Participation in village meetings may be easy to manipulate at no cost, however. We provide an additional measure of participation that focuses on the process in which the committees were selected. 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 (normalized) composite measure, combining these four indicators, in Table 2 illustrates ("*Participatory Selection Methods*"), we find no evidence that participation in *Tuungane* leads to greater adoption of participatory processes in the selection of the committee or projects.

Last, if the average villager is more likely to effectively participate, we should expect RAPID committees to have a broader representation of the population. We implemented an additional measure of participation: the composition of the RAPID committee ("*Committee Composition*"). 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.[11]

4.2 Accountability

A set of accountability measures are used to assess whether exposure to good governance practices led to adoption of similar mechanisms.

First we combined survey and focus group data to construct a measure of what kinds of accountability mechanisms were organized to oversee committees. We find that in the majority of villages no mechanisms had been put in place to oversee the use of RAPID

[10] We find similar results for dominance of chief and elderly.

[11] 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.

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 had been required to report its actions to the community as a whole. As the composite measure in Table 2 indicates ("*Accountability Mechanisms*"), *Tuungane* did not lead to a greater propensity to put accountability mechanisms into place.[12]

Second, during during Step D, we asked ten randomly selected 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 ("*Private Complaints*") suggest that levels of complaint are no higher in *Tuungane* areas than in control. Our preanalysis plan called for a comparison of complaints between treatment and control but did not propose examining differences in behavior conditional on there being something to complain about. We subsequently added this analysis as it more closely approximates the concern of interest. In this one unregistered test ("*Private complaints — Mismanagement*"), we find the strongest evidence for a *Tuungane* effect, with a more than doubling of the responsiveness of complaints to fund mismanagement.[13]

4.3 Efficiency

A set of efficiency measures are used to capture changes in capacity and effort by elites and the broader population as well as the rate of information transmission in villages.

Measures of accounting quality were based on an accounting form that was distributed as project funds were transferred (Step C). RAPID committee we expected to indicate the total amount made available for the project (out of $1000), and to keep track of expenditures made on this form. We use the presence of this form, and information on whether it was completed as an indicator of efficiency implementation. We found that on average, in 82% of the villages, committees had their 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 ("*Quality of Accounting*") 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

[12] 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 random villagers).

[13] 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.

of accounting.

We generate a second behavioral measure of the extent to which the community can function efficiently outside of the RAPID process by examining information transmission among villagers. For this we delivered public health information on hygiene and diarrhea to a random sample of villagers in the first visit. We then returned to these villages (step D) and interviewed the same villagers as well as a random sample of villagers who were not yet visited. We compare the answers provided by these additional villagers in a health information test to the answers provided by villagers in another 396 randomly selected villages in which no information was delivered but the tests were implemented simultaneously. This allows us to assess the rate of information spillovers within villages, and compare it between treatment and control. Overall we found that those living in villages where we distributed the information to other people score ten points higher on a set of questions related to the public health information we provided. This result indicates that information spreads. As Table 2 ("*Information Transmission*")) shows however, we see no evidence that *Tuungane* had an 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.

4.4 Transparency

Two further measures of information transmission are used to assess whether the principles of transparency emphasized by the intervention were subsequently adopted.

To measure information delivery by the elite to the villagers, the enumerator teams informed the community in the first meeting (Step A) that $900 *or more* will be made available. In all communities, however, the self-identified committee received $1000 after one week during Step C. We are able to measure information delivery by interviewing a random sample of villagers about the amount of the RAPID grant to see whether they report the figure told to them by us in Step A, or rather the true amount, as known by committee members. Table 2 (*Knowledge of project amount*) shows that, on average, 39% of all respondents (and 56% of those respondents that gave an answer) report the correct answer of $1,000. We find no evidence that there is a difference between treatment and control communities however.

To assess the willingness and ability of randomly selected villagers to obtain relevant information about the management of public resources, we went a step further asking a sample of 1,406 respondents to seek out information on fund use in their communities. From row 12 (*Willingness to seek information*) we see that 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 exercise 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). This suggests that accessing basic financial information is challenging. There is no significant difference between treatment and control however.

4.5 Capture

The final and perhaps most important set of measures assess whether overall changes in behavior led to more equitable distributions of benefits.

As a first key measure we use the amount of the $1,000 grant that auditors were unable to account for. To construct this estimate, the auditors visited nearby markets to verify measures of price and quantity as listed in the accounting form. Table 4 indicates the auditors' check-list. Table 2 (*(Fewer) Financial irregularities)*) presents the results: on average $852 of the $1,000 could be verified by the teams. There is no significant difference between *Tuungane* and control communities. This suggests little difference in fraudulant behavior across groups, though this does not itself mean that resources that could be accounted for were used well by either or both groups.

We next look at the distribution of economic benefits from the RAPID project. As a direct measure of capture of economic benefits, we implemented a survey to a random sample of claimed beneficiaries of the project and observed whether they actually existed, and the value of the benefits delivered. We use the standard deviation of benefits as a measure of the inequality of benefits distributed. 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. We compute the 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 ("*Inequality of benefits*"), indistinguishable from *Tuungane* communities

Finally, we provide results from a behavioral measure that captures the extent to which actual decisions disproportionately reflect the preferences of the village chiefs sorts of villagers. our measure compares the predictive power of the chief's preferences to those of a random sample of other villagers. We find ("*Dominance of preferences*") that in all areas 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. 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. Chiefly dominance is around one point lower in *Tuungane* areas. We thus find evidence of chief dominance in all areas but little evidence that the *Tuungane* program reduces the strength of this dominance.

Overall we find almost no evidence for causal effects. This may be because there are no effects. But it may also because of our case selection, or because of biases in measurement or estimation. We turn to examine these possibilities next.

5 Alternative explanations

We first consider case selection and then consider two possible biases. We begin by assessing whether spillovers to neighboring areas are present, as these could lead to a downward bias in estimated treatment effects. We then assess whether a countervailing backlash mechanisms is in operation that could mask other positive effects and whether weak results could be due to differential desirability bias. In addition, in the online appendix (I) we discuss the possibility that results reflect a bias due to a short term elite response to the intervention, and then explore concerns related to data missingness, compliance, treatment heterogeneity and specification biases (J).

5.1 A bad case?

One possible explanation for weak evidence of effects is that this was simply a weak intervention and not typical of the kind that researchers or policy makers expect to generate strong effects. Did development funders and implementers supporting this project expect that it would produce strong effects? To find out, and prior to launching our endline data collection, 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. The survey simply elicited beliefs regarding likely impacts on each of the outcomes in different categories. It was not incentivized. The responses showed 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 though they were more optimistic about effects on participation and considerably more skeptical that transitional leaders would become more accountable (most researchers reported that they would not).

Access to this information is valuable for the simple reason that it was formed prior to data gathering. If the weakness of the intervention seem 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 models.

These beliefs reflect confidence that CDD is an effective model but also that this particular project was well implemented. Our data using interviews with members of the population, village chiefs and VDC members in all *Tuungane* villages also support the view that the project was well implemented. From this data we find that 62% of the population in *Tuungane* villages knew about *Tuungane* and 39% of those knew who implemented it. Furthermore, 76% of VDC members and 48% of village chiefs were able to guess the right size of the grant, although only 22% of the general population guessed the grant amount.

Finally, we also recorded attendance at project meetings. We find that 35% of the population reported attending some meetings associated with *Tuungane*. More than half of the chiefs interviewed reported attending some meetings and 84% of VDC members reported attendance. The median participant villager attended 2 meetings, with the top 25% claiming to have attended more than 4. The median participant chief reported attending four meetings while the top 25% attended 7 or more; the equivalent numbers for the VDC members are 9 and 15 meetings.

The overall knowledge and participation in the project was therefore considerable, and thus the lack of exposure to the democratization components is unlikely to explain our null results.

5.2 Spillovers

If *Tuungane* produced positive effects beyond treatment communities, our results could be the result of positive spillovers.

We use a design based strategy to assess the presence of spillovers.[14] That is, we define an "x-km indirect effect" as the effect of being within x kilometers of a *Tuungane* village that is part of another cluster of villages.[15] The propensity of being exposed to such a treatment

[14] See Gerber and Green (2012), Ch 8, 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.

[15] 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 location 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).

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. We make use of the random assignment to recover these propensities, since they are determined 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. We then generate estimates of treatment effects 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).[16]

We conduct our analysis for both a 5km radius spillover treatment and a 20km radius spillover treatment. We highlight (and illustrate in Figure 4 in Supplementary materials) that when we examine different conceptualizations of the treatment effect we *simultaneously alter our samples*. The intuition is that a unit in a block 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.[17] Finally, we note that while our estimates 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).

The results (presented in Table 6 in online appendix J) shows no evidence that exposure

[16] 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. See also: Barrios et al. (2012) .

[17] 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).

to *Tuungane* had spillover effects on local power structures (or direct effects, once we take account of possible spillovers). Our estimates are supported by the following facts. First, the absence to detect a main effect as a result of spillovers into control villages is consistent only with unusually large spillovers of 100%. Second, randomization was implemented at the level of clusters of VDCs, hence 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, populations in control areas report low levels of knowledge about *Tuungane*.

5.3 Differential Desirability Bias

A final possible bias we consider is that control communities, expecting future conditional aid disbursement, and having not received the infrastructure funds, may have managed the cash in a more democratic manner in order to please future donors. This could be the case if the RAPID cash delivery project was perceived as linked to the donor community, despite our best efforts to uncouple them.[18] We use a small experiment embedded in our endline survey to shed light on this possibility.

To assess directly whether villagers strategically displayed behaviors aligned with their beliefs about expectations of development donors, we introduced a survey variation in which 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 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.[19]

We find very strong evidence for a social desirability bias: in both groups individuals are

[18] 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.

[19] More precisely, we gave both prompts to *all* respondents but randomized the order of the prompts. Though not exploited here this allows us to generate a within person measure as well as consistency bias. The results here use only the first prompt however which provides the cleanest results.

19 - 20 percentage points more likely to answer yes following a positive prompt (standard error: ca 0.025 / 0.07) (full results are provided in Table 5 in Supplementary Materials.). Importantly, however, we do not find evidence that this bias is affected by exposure to *Tuungane*. The difference between the two groups is small but is stronger (though not significantly so) for the *Tuungane* group. Therefore, social desirability bias among control communities is unlikely to drive our result.

6 Conclusion

We replicate and extend past studies of the effects of external efforts to export institutional innovations to developing countries. By engaging in replication, our research takes seriously the goal of cross study cumulation. We do so here in an area where a line of theory and common belief among development practitioners supports a view that policy interventions that expose populations to new democratic practices will lead to adoption and persistence. This idea underlies a large class of development aid projects including many of the largest interventions in post-conflict areas (Mansuri and Rao, 2013). The logic underlies many other interventions also in the areas of development and governance that seek to change institutions while leaving structural features, such as income distribution or property rights, intact.

Yet, using strong behavioral measures and a large sample, our study adds to a growing body of findings that suggests there is something wrong with this rosy view. We examine many possible explanations for our negative results and these investigations give confidence that we are not reporting a false negative.

Our findings contribute to the literature on the political economy of development and to development practice. First, our findings suggest that institutional change is unlikely to spontaneously emerge as a result of demonstration effects and exposure to new institutions. This has implications for scholarship studying long-run development and institutional change (Akyeampoing et al., 2014). Second, our findings suggest the need to rethink the strategies employed by governments in developing countries and donors to improve institutions. Our findings suggest that current donor-driven approaches to render decision-making more inclusive by short/medium-term interventions which do not change the economic fundamentals may be flawed: when the elite is likely to have vested economic interests, institutional change may itself require a change in the allocation of economic power.

In closing, we highlight one caveat and one clarification.

The caveat relates to one interpretation of our null effect that we cannot completely rule out. The basic logic of positive exogenous institutional change is that external action helps shift populations from one equilibrium to a better one. This presupposes that these populations are in a bad equilibrium in the first place. This is a common view for the type of problem we are examining. Scholars frequently adopt a chiefs-as-despots model (Acemoglu, Reed and Robinson, 2014) and view rural institutions as captured by traditional elites. Moreover CDD programs are often motivated by a claim that past conflict destroys social cohesion and governance capacity. This view is partly supported: we uncover financial irregularities affecting 15% of funds across areas, a domination of chiefly preferences across areas, and less than 50% knowledge of project amounts. Moreover we found that very few respondents felt they had a right to seek out information on fund usage in their communities. However, we also found higher baseline levels of general participation, public information, and perceived legitimacy of existing decision making mechanisms across these communities than expected. This suggests that the despotic view of rural institutions is potentially overdrawn (see also Logan (2013)). Without a complete mapping of the set of equilibriums that might exist we cannot be sure that the lack of change we see is due to the resilience of suboptimal equilibriums or the absence of optimal alternatives.

The clarification relates to the extent of the implications of our claim. We take aim at a large target: that external interventions can have a large effect on the behavior of groups without any change to fundamentals. Large targets are easily hit but not easily taken down, and that is the case here. We cannot conclude from this study that the institutional claim is incorrect. Indeed, when phrased as a possibility claim, it resists falsification. The implications of our findings relate more to the application of the claim than the claim itself: the institutional logic provides grounds for optimism that short term interventions may produce changes in expectations that produce large changes in behavior without any accompanying changes in fundamentals. Our results, and others, speak against that inference in cases where development actors have placed great confidence in it. Future work should assess whether the limited effects arise because in contexts with multiple equilibrium, changing expectations is harder than development actors have supposed, or because in these contexts, variation in outcomes depends more on changes to fundamentals and less on equilibrium selection.

References

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

Acemoglu, Daron, Tristan Reed and James A Robinson. 2014. "Chiefs: Elite Control of Civil Society and Economic Development in Sierra Leone." *Journal of Political Economy* 122(2):319–368.

Akyeampoing, Emmanuel, Robert Bates, Nunn Nathan and James Robinson. 2014. *Africa's Development in Historical Perspective*. Cambridge: Cambridge University Press.

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

Avdeenko, Alexandra and Michael J Gilligan. N.d. "International Interventions to Build Social Capital." *American Political Science Review*. Forthcoming.

Barrios, Thomas, Rebecca Diamond, Guido W. Imbens and Michal Kolesár. 2012. "Clustering, Spatial Correlations and Randomization Inference." *Journal of the American Statistical Association* 107(498):578–591.

Barron, Patrick, Macartan Humphreys, Laura Paler and Jeremy M. Weinstein. 2009. "Community Based Reintegration in Aceh." *Indonesian Social Development Papers* 12.
URL: *http://tinyurl.com/q2vyend*

Beath, Andrew, Fotini Christia and Ruben Enikolopov. 2013. "Do Elected Councils Improve Governance? Experimental Evidence on Local Institutions in Afghanistan." *World Bank Policy Research Working Paper Series* (6510).
URL: *http://elibrary.worldbank.org/doi/pdf/10.1596/1813-9450-6510*

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

Bowers, Jake and Mark M Fredrickson. 2013. "Reasoning about Interference Between Units." *Political Analysis* 21:97–124.

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

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

Fearon, James D., Macartan Humphreys and Jeremy M. Weinstein. 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, Ronald A. 1935. *The Design of Experiments*. London: Oliver and Boyd.

Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York City: W.W. Norton.

Glaeser, Edward L, Rafael La Porta, Florencio Lopez-de Silanes and Andrei Shleifer. 2004. "Do institutions cause growth?" *Journal of economic Growth* 9(3):271–303.

Greif, Avner and David D. Laitin. 2004. "A Theory of Endogenous Institutional Change." *American Political Science Review* 98(4):633–652.

Grossman, Herschel I and Minseong Kim. 1995. "Swords or plowshares? A theory of the security of claims to property." *Journal of Political Economy* pp. 1275–1288.

Hamilton, William D. 1964. "The Genetical Evolution of Social Behavior." *Journal of Theoretical Biology* 7(1):1–16.

Herbst, Jeffrey. 2014. *States and power in Africa*. Princeton University Press.

Humphreys, Macartan, Raul Sanchez de la Sierra and Peter Van der Windt. 2011. *Tuungane I: Outcomes and Data Sources*.

Humphreys, Macartan, Raul Sanchez de la Sierra and Peter Van der Windt. 2012. Social and Economic Impacts of Tuungane. Technical report.

IRC. 2007. *Mortality in the Democratic Republic of Congo*. IRC.
URL: *http://www.rescue.org/resource-file/irc-congo-mortality-survey-2007*

King, Elisabeth and Cyrus D. Samii. 2014. "Fast-Track Institution Building in Conflict-Affected Countries? Insights from Recent Field Experiments." *World Development* 64(1):740–760.

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

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

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

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

Mehta, Judith, Chris Starmer and Robert Sugden. 1992. An Experimental Investigation of Focal Points in Coordination and Bargaining: Some Preliminary Results. In *Decision Making under Risk and Uncertainty*. Springer pp. 211–219.

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

Nunn, Nathan and Nancy Qian. 2014. "US Food Aid and Civil Conflict." *American Economic Review* 104(6):1630–1666.

Rodrik, Dani, Arvind Subramanian and Francesco Trebbi. 2004. "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development." *Journal of Economic Growth* 9(2):131–165.

Ross, Michael L. 2001. "Does Oil Hinder Democracy?" *World Politics* 53(3):325–361.

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

Sachs, Jeffrey D. 2005. *Investing in Development: A Practical Plan to Achieve the Millennium Development Goals*. UN Millennium Project Earthscan.

Shayo, Moses. 2009. "A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution." *American Political Science Review* 103(02):147–174.

Shepsle, Kenneth A. 2006. "Rational Choice Institutionalism." *The Oxford Handbook of Political Institutions* pp. 23–38.

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

Weingast, Barry R. 1995. "The Economic Role of Political Institutions: Market-Preserving Federalism and Economic Development." *Journal of Law, Economics, & Organization* 11(1):1–31.

Wong, Susan. 2012. "What Have Been the Impacts of World Bank Community-Driven Development Programs?" *World Bank, Washington, DC* .

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: Princeton University Press.

Supplementary Material

A Institutional Logics

Consider a simple game in which two players, Strong (S) and Weak (W), can each decide in each of an infinite number of periods whether to produce using a default technology (D) or a cooperative technology (C). Say each period decision resembles a prisoner's dilemma. If both use the cooperative technology they produce output worth 1 unit. If both stay with the default technology their yield is $d_j = .5$ for $j \in \{S, W\}$. If one uses the default technology while the other attempts to use the cooperative technology on her own the first receives free-rider yield $f_j \in (1, 2)$ while the second receives 0. In addition, players can make cash transfers to each other (assuming utility is linear in income, we treat utility as transferable).

Baseline equilibrium. With sufficient patience, the following is a subgame perfect equilibrium of this game: both players cooperate every period, each producing .5 units of value more than they would over the returns using the default technology. Player W then transfers .4 units of value to player S, and players end the round with payoffs of .6 for W and 1.4 for S. If in any period a player plays D or the appropriate transfer is not made, then all players play D in every subsequent period.

In this equilibrium S extracts 80% of what W produced over and above what she would have gained had they both played D. Following Greif and Laitin (2004) this equilibrium *is* the institution, it is sustained by equilibrium expectations of players that cooperation will only be sustained if W makes large transfers to S. In this case we might think of the political part of the institution as the 80% tax rate imposed on W.

Suppose now a third party views this equilibrium as exploitative and seeks to change outcomes. Consider two strategies they might employ.

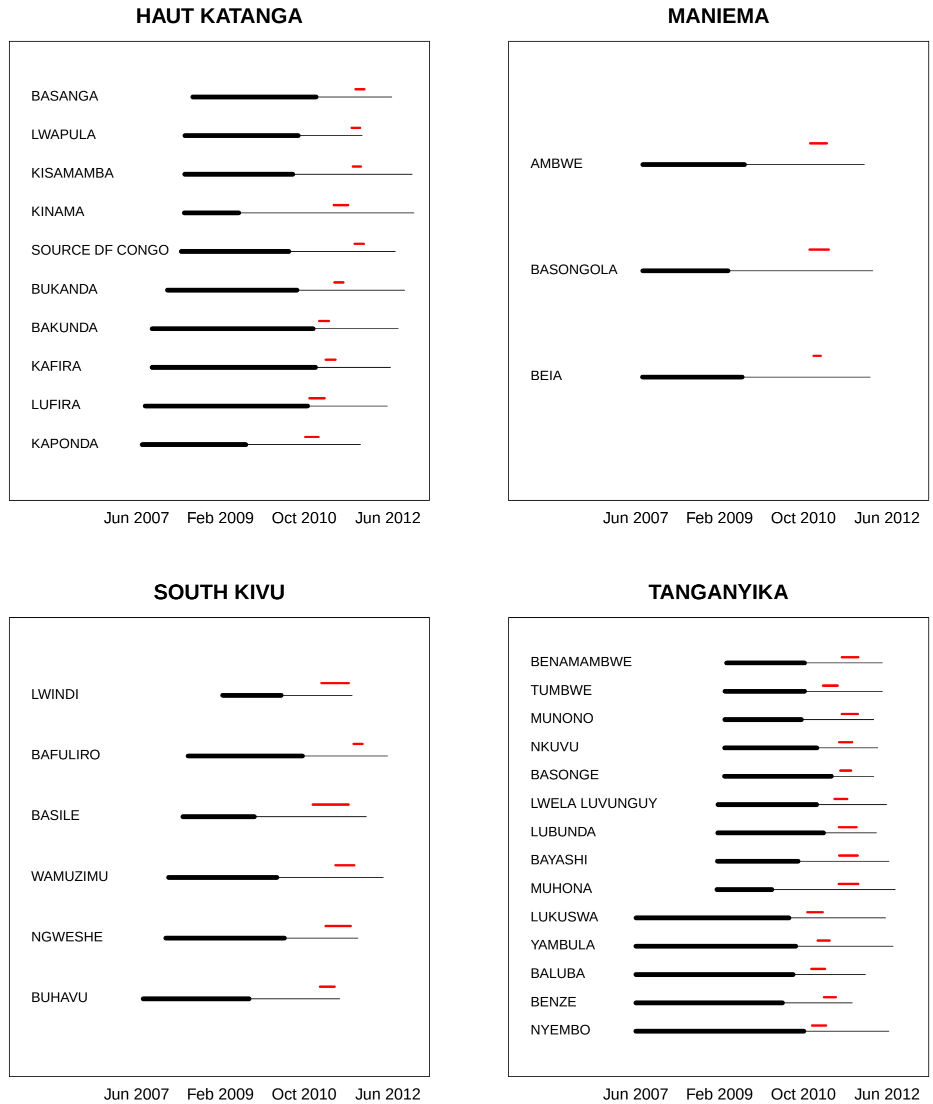
Strategy A**.** The first strategy seeks to improve the lot of W by changing the equilibrium. Leaving the game intact, the third party proposes that the surplus be divided more equally, perhaps proposing that W only transfers half as much each period to R, leaving W and S with 1.2 and 0.8 respectively. The strategy is motivated by the observation that a 40% tax regime (on surplus) can also be sustained in equilibrium and so if players adopt the right expectations the new transfers will be self-enforcing. This intervention is a purely institutional intervention: it focuses on expectations and leaves the underlying game unchanged.

Strategy B**.** Consider now a second intervention in which the third party guarantees W a return of $d_W = 0.75$ instead of $d_W = 0.5$ in the event of cooperation failure. This is a structural change and has a real effect on W's bargaining position. It means that W can now do better playing D in all periods and giving up cooperation with S. Both will still do better under some cooperative arrangement however. Say in the event of cooperation, S continued to extract 80% of W's surplus. Then she would now force a transfer of $.25 \times .8 = .2$ and so W would be left with 0.8.

Strategy B produces the same outcome $(0.8, 1.2)$ as achieved by Strategy A but does so without requiring a change in the approach used by the players to divide the surplus. Moreover the behaviors *on the equilibrium path* following the two interventions are the same — both players play C, each earns 1 unit and W transfers .2 units to S. The effect of Strategy B however is not due to changes in the equilibrium selected but to a change in the underlying game (albeit one that matters only off the equilibrium path).

B Timing of Intervention and Measurement

Figure 2: Timeline of Implementation



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

C Balance *Tuungane* Treatment and Control

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 pretreatment 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 Humphreys, Sanchez de la Sierra and Van der Windt (2011). Table 3 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 3: Balance

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
Sex 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 (Humphreys, Sanchez de la Sierra and Van der Windt (2011)) 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 3 does not change our results (not reported).

D RAPID Script Meeting Step A

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 community's 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?"

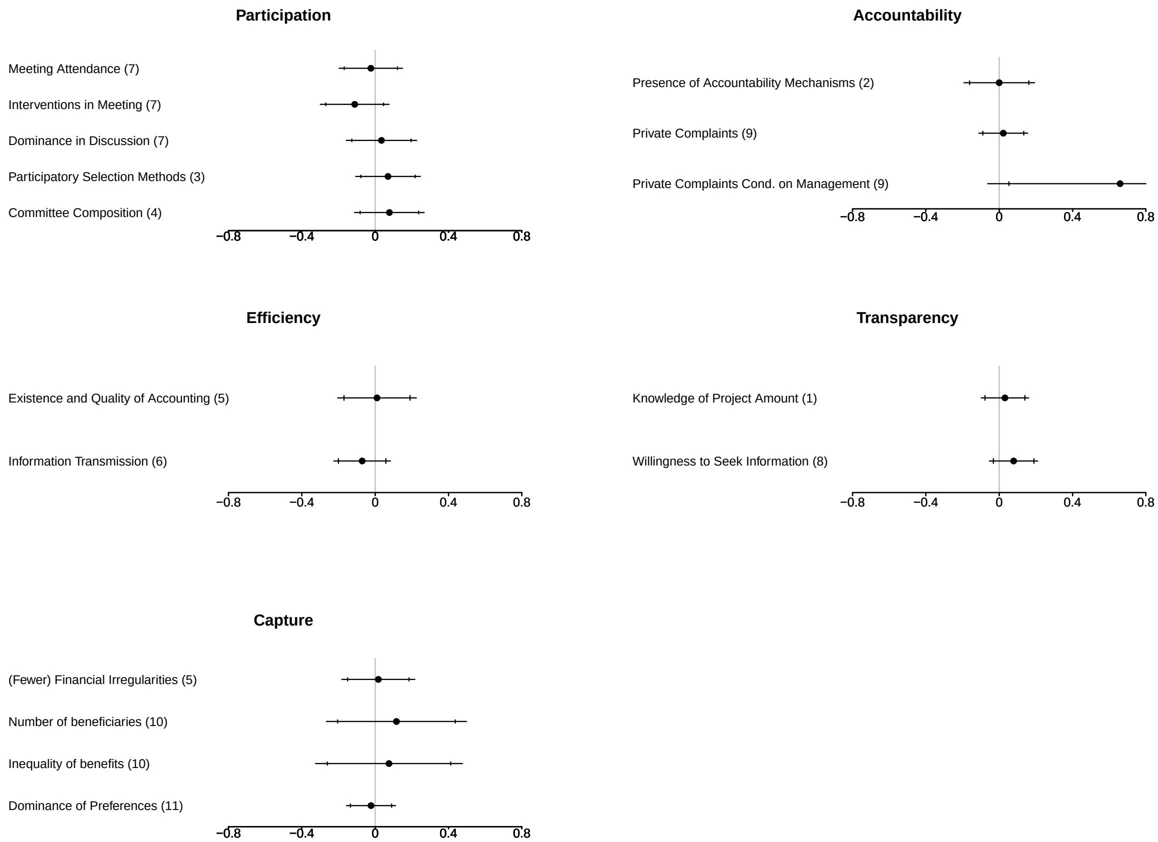
E Auditor Check-list

Table 4: The Auditor check-list to detect fraud stemming from mis-reporting of purchases

Survey item	Strategy of the committee	Action necessary by the auditor
DA 53	**The committee may inflate each price**	**The committee may try to conceal the relationship with the auditor must verify the actual price at time of purchase and the amount indicated:**
		Talk to the mothers of the village (often it is the mothers who know more about market prices)
		If possible, go check the market (or in lieu of purchase), talk to the vendor, other vendors to confirm these mothers and
DA 54	**The committee may falsify receipts**	**The auditor shall verify the receipts obtained:**
		Check with the committee (the date of receipt, the seal on the receipt, writing of the receipt (which should be different from writing about FC)
		Check receipts with others that the committee, who can tell you if any, and if these suppliers are credible or friends of the committee
DA 55	**The committee can make a plot with suppliers to lie to you**	**The auditor should verify that suppliers do not lie:**
		Check if the suppliers change their speech in front of other people
		If it is a market, ask other suppliers
		Ask other people present at the location of suppliers
DA 56	**The committee may use the property as to deceive you: he can buy goods of inferior quality**	**The auditor should verify the quality of what was purchased:**
		Inform yourself about the different types of existing quality in the middle and prices (either mothers or market)
		Inquire about how to check the quality of the objects in question
DA 57	**The committee may add false transportation costs**	**The auditor should verify the true costs of transport:**
		Make sure the committee has directed the transport indicated by asking several people in the village -
		Learn about the real cost of transport in question in the village —
DA 58	**The committee can show the property or facilities that already exist and claim they were made by the RAPID project**	**The auditor should verify that the facilities (or goods) did not already exist:**
		Installation: check its age and condition. Ask villagers the date of construction of the facility
		Property distribution: when you are with the beneficiaries, ask to see distributed objects and ask the household how long the property exists there in the household and where is it from (the committee may also distribute goods he had already stored somewhere)
DA 59	**The Committee can claim that the project was completed while he has not been**	**The auditor should verify the existence of the project:**
		Facility: Field
		Distribution: With the 10 beneficiaries
DA 60	**Use local materials and pretend they were purchased elsewhere**	**The auditor should verify that the materials have been purchased elsewhere by going to check with the seller and talking to the population**
DA 61	**The committee may use the seasonal variation in prices**	**The auditor must ensure the price verification that this is the price at the time of purchase, not today**
DA 62	**The committee may try to conceal the relationship with suppliers**	**The auditor should verify the relationship:**
		Calling on the suppliers themselves
		Talking with people in the village Suppliers

F Main Results: Figure

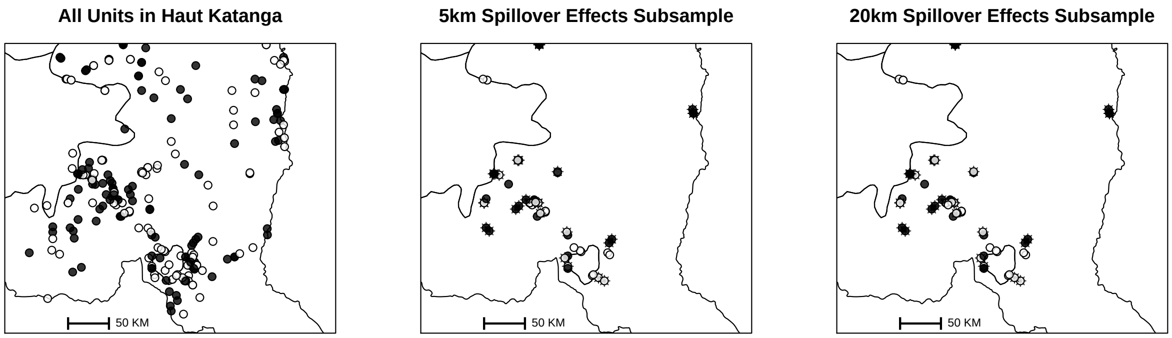
Figure 3: Main Results



Notes: Treatment effects reported by family. All analyses employ propensity score weights and clustering of standard errors at the level of randomization clusters. See table version of this figure in the main text. For this figure all estimates are reported in units of standard deviations of the outcome in the control group.

G Spillovers

Figure 4: 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 sub-sample 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.

H Differential Desirability

I Elite Backlash against loss in power

Since village chiefs were actively excluded from the *Tuungane* program, they might have had incentives to seek compensation during RAPID. In this case the null result could reflect unusually strong incentives for traditional leaders to engage in capture in treatment groups, coupled with strong restraints induced by bottom up pressures following the intervention. In Section 4 we found that the implemented projects (obtained during Step D) coincide better with the stated preferences (taken during Step A before the village meeting) of the chief than those of the villagers. We interpreted this as possible evidence for chief dominance. To explore whether the chief captured the RAPID process, and especially so in *Tuungane* areas, we investigate whether in *Tuungane* areas members of the RAPID committee are more closely related to the village chief.

To measure network proximity, we collected detailed friendship and kinship data among randomly selected villagers, which includes their relationship with all committee members

Table 5: 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?"

as well as with the chief. We also collected detailed friendship and kinship data among all committee members, which includes their relationship with the village chief. We then create a measure of family connection to the committee using the Hamilton index.[20] We find neither the population nor the chief are closely related to RAPID committee members. The average score on our index for the population is 3.49%, while the score for the village chief is 4.45%. This difference is statistically significant, and amounts to the chief adding a first cousin to the RAPID committee.[21] We find no difference in this kinship proximity between *Tuungane* and non-*Tuungane* areas however.

Other measures also confirm that the chief did not disproportionately dominate procedures in *Tuungane* areas. During Step B our enumeration team led focus groups with ordinary villagers to learn whether the process of committee and project selection was electoral, by lottery, by consensus, imposed by the chief, by elders, other or unknown. Very few people (around 5%) find that the chief imposed project selection or committee member selection and equally so in treatment and control areas. Finally, during Step D we directly ask individuals whether the RAPID committee was controlled by the chief. Around 26% of the 2,514 individuals answer in the affirmative. However, from a menu of thirteen complaints

[20]The Hamilton index measures the biologic relatedness between two individuals: for a parent-offspring or full sibling relationship this index is 50%, for an aunt/uncle or nephew/niece relationship this is 25%, etc. See: Hamilton (1964). Applied to the group, if for example, two members of the RAPID committee, out of the five, are children of the chief and one is a nephew, the chief's Hamilton score is 25%.

[21]Note, however that in almost 63% of the villages have no relationship at all to the chief.

less than 5% of the respondent find this to be the most important complaint. Moreover, there are no differences in reporting across treatment and control areas. We thus conclude that the null result reported in this paper does not reflect chiefs' response to *Tuungane*.

J Robustness

In the text we discuss concerns related to spillovers and to social desirability biases. Here we describe issues related to attrition and data missingness, noncompliance, treatment heterogeneity, and specification sensitivity.

J.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).[22] 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.

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

[22] 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.

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).[23] 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.[24]

J.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

[23] 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.

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

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 6) are similar.

J.3 Treatment Heterogeneity

As seen in Figure 2 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.

J.4 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 village 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 6) are robust to these different specifications.

Table 6: Robustness

Concept	Measure (#)	Base	Alt. Treat.	Spillover Effects D(5km)	I(5km)	D(20km)	I(20km)	Alt. Specifications Village	Bins	Prop.
Participation	Meeting Att. (7)	-1.98	-3.98	-5.38	22.19	0.55	-5.62	-1.98	-1.24	-1.56
		(0.79)	(0.59)	(0.824)	(0.489)	(0.467)	(0.97)	(0.79)	(0.83)	(0.84)
	Interventions (7)	-0.49	-0.19	-0.44	-0.05	-0.40	-0.25	-0.49	-0.37	-0.42
		(0.25)	(0.66)	(0.707)	(0.246)	(0.827)	(0.371)	(0.25)	(0.31)	(0.32)
	Dominance (7)	0.52	-0.33	0.49	-1.21	0.05	0.04	0.52	0.52	0.74
		(0.73)	(0.82)	(0.362)	(0.321)	(0.482)	(0.216)	(0.73)	(0.68)	(0.62)
	Selection Meth. (3)	0.07	0.05	0.12	0.28	-0.03	0.42	0.07	0.07	0.08
		(0.45)	(0.59)	(0.262)	(0.305)	(0.646)	(0.302)	(0.45)	(0.30)	(0.40)
	Committee Comp. (4)	0.08	0.10	0.18	0.05	0.14	0.24	0.08	0.09	0.07
		(0.42)	0.1	(0.08)	(0.869)	(0.08)	(0.171)	(0.42)	(0.19)	(0.44)
Accountability	Accountability Mech. (2)	0.00	-0.02	0.00	0.09	-0.09	-0.18	0.00	0.00	-0.02
		(0.97)	(0.83)	(0.47)	(0.37)	(0.788)	(0.849)	(0.97)	(1.00)	(0.85)
	Complaints (9)	0.02	-0.01	0.13	0.24	-0.01	0.31	0.01	0.01	0.00
		(0.75)	(0.89)	(0.129)	(0.65)	(0.582)	(0.566)	(0.90)	(0.89)	(0.99)
	Complaints Cond. (9)	0.68	0.71	0.56	0.24	0.56	-0.17	0.82	0.57	0.77
		(0.07)	(0.06)	(0.11)	(0.55)	(0.14)	(0.81)	(0.02)	(0.05)	(0.03)
Efficiency	Accounting (5)	0.01	-0.05	-0.12	-0.37	0.00	-0.60	0.01	0.01	0.01
		(0.90)	(0.63)	(0.695)	(0.602)	(0.491)	(0.608)	(0.90)	(0.88)	(0.95)
	Info Transm. (6)	-1.41	0.08	-2.41	-1.17	0.20	-0.34	-0.61	-0.74	-0.98
		(0.37)	(0.96)	(0.22)	(0.53)	(0.90)	(0.88)	(0.67)	(0.61)	(0.50)
Transparency	Knowledge (1)	1.52	-0.73	-2.95	-9.73	0.30	-11.98	1.27	1.44	1.69
		(0.64)	(0.80)	(0.696)	(0.308)	(0.466)	(0.415)	(0.66)	(0.52)	(0.55)
	Seek Info (8)	3.84	4.01	5.82	1.90	0.51	12.31	2.19	1.67	1.35
		(0.24)	(0.16)	(0.103)	(0.78)	(0.421)	(0.114)	(0.48)	(0.45)	(0.66)
Capture	Fin. Irregularities (5)	3.52	-13.94	-27.51	-15.41	0.84	-46.85	3.52	5.53	1.19
		(0.87)	(0.50)	(0.836)	(0.293)	(0.486)	(0.458)	(0.87)	(0.76)	(0.96)
	Beneficiaries (10)	3.28	6.41	4.71	-2.91	1.07	11.77	3.09	-1.98	3.05
		(0.55)	(0.21)	(0.344)	(0.906)	(NA)	(NA)	(0.54)	(0.63)	(0.54)
	Inequality (10)	0.56	-0.14	-0.67	1.98	-0.70	1.61	-0.15	0.58	-0.11
		(0.71)	(0.92)	(0.706)	(0.25)	(NA)	(NA)	(0.92)	(0.65)	(0.94)
	Dominance of Pref. (11)	-0.01	0.00	-0.03	-0.01	-0.05	0.09	-0.01	-0.01	-0.02
		(0.73)	(0.92)	(0.28)	(0.85)	(0.30)	(0.20)	(0.66)	(0.64)	(0.58)

Notes: 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 level of randomization clusters 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 village 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).

J.5 Heterogeneous Effects

Our results may mask positive effects for population sub-groups. In particular, it is plausible that democratization was already advanced in most areas, and only in areas subject to a lot of capture by local elites did the program have an effect. We rule out this alternative explanation by identifying pre-existing levels of capture and estimating the heterogeneous effects of *Tuungane* by this pre-treatment characteristic.

To measure pre-treatment capture, we use three different indicators, based on data from our surveys with village chiefs. First, we look for the degree of competition by which current chiefs acquired their position. Absence of competition to local chiefs has been described to be a major driver of chiefs' capture of communities and civil society (Acemoglu, Reed and Robinson, 2014). To measure the degree of competition, we identify villages where the village chief inherited his position from his father. While 37% of chiefs' positions were inherited, 10% were chosen by elders, 25% were chosen by the local Mwami (traditional head of a large territory), 14% were chosen by other chiefs and 14% were chosen by elections. Second, we construct indicators of community mobilization. We identify villages without village association or committee before the start of *Tuungane*; this leaves us with 170 of the 358 villages. Third, we identify villages that had no classrooms in July 2006. This is the case for 194 of the 358 RAPID villages. We believe that areas where capture is effective will have lower public goods provision, especially for basic services such as education.

Table 7 presents the results from the subgroup analysis. We find very little evidence of heterogeneous effects in favor of positive effects among captured communities. In addition to a few positives, an equal number of coefficients are negative: interventions in meetings and information transmission for those villages with no projects in 2007, and interventions in meetings and dominance of preferences for those villages where the chief's father inherited his position.

Table 7: Heterogeneous effects of grassroots democratization, by initial institutions

					Subgroup:								
		Main			Projects			Committees			Inherited		
Outcome	Measure (Measure Number)	Effect	(se)	N	Effect	(se)	N	Effect	(se)	N	Effect	(se)	N
Participation	Meeting Attendance (1)	-1.98	(7.40)	455	-5.38	(11.14)	187	-0.12	(8.93)	269	-9.80	(15.31)	129
	Interventions in Meeting (1)	-0.49	(0.42)	457	-1.13	(0.67)	189	-0.47	(0.56)	271	-1.49	(0.75)	129
	Dominance in Discussion (1)	0.52	(1.49)	457	-1.08	(2.44)	189	-0.70	(1.91)	271	1.81	(2.89)	129
	Part. Selection Methods (2)	0.07	(0.09)	451	0.29	(0.15)	188	-0.01	(0.12)	269	0.11	(0.17)	129
	Committee Composition (3)	0.08	(0.10)	452	0.04	(0.15)	190	0.04	(0.12)	271	-0.15	(0.19)	128
Accountability	Presence of Acc. Mechanisms (4)	0.00	(0.10)	414	-0.13	(0.15)	194	0.08	(0.14)	225	-0.13	(0.16)	132
	Private Complaints (5)	0.02	(0.07)	3647	0.05	(0.10)	1754	0.08	(0.10)	1939	-0.04	(0.12)	1185
	Private Complaints Cond. (6)	0.68	(0.38)	3502	0.51	(0.46)	1661	1.39	(0.67)	1855	0.00	(0.47)	1148
Efficiency	Existence and Quality of Acc. (6)	0.01	(0.11)	399	-0.07	(0.17)	185	0.01	(0.15)	216	0.15	(0.18)	129
	Information Transmission (7)	-1.41	(1.56)	3800	-3.34	(2.46)	1935	-1.46	(2.25)	2106	-1.50	(2.92)	1187
Transparency	Knowledge of Project Amount (8)	1.52	(3.21)	3685	3.68	(4.70)	1760	1.64	(4.46)	1965	1.72	(5.60)	1198
	Willingness to Seek Information (9)	3.84	(3.28)	1406	7.61	(4.36)	723	3.74	(4.23)	760	1.48	(5.82)	433
Capture	(Fewer) Financial Irregularities (10)	3.52	(20.74)	394	-6.43	(29.43)	182	10.92	(25.44)	214	-4.05	(41.66)	128
	Number of beneficiaries (11)	3.28	(5.52)	154	16.19	(7.83)	76	1.08	(7.66)	80	-3.71	(10.63)	53
	Inequality of benefits (11)	0.56	(1.52)	127	1.78	(2.20)	64	0.69	(2.14)	71	2.78	(2.65)	45
	Dominance of Preferences (12)	-0.01	(0.03)	2666	0.00	(0.04)	1098	-0.01	(0.04)	1583	-0.07	(0.05)	752

Notes: This table replicates Table 2 for three subgroups. The column under 'Main' are the results as presented in Table 2. Results under 'Projects' are the results for villages without school rooms in 2007, "Committees" to those without any village committee or association in 2007, and 'Inherited' to those villages where the chief's father's position was inherited.

K 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. All of our analysis were based on hypothesis 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– circulated to colleagues and posted online (this was prior to the existence of a social science registry to which we could post the analysis plan). The analysis presented here differs from those described in the analysis plan in four ways.

First, we focus here on a subset of tests, specifically we focus on the *behavioral* tests of *governance* effects. All other tests have however been implemented and are available in supplementary material (see Humphreys, Sanchez de la Sierra and Van der Windt (2012)). To clarify the relation between the hypotheses listed in design documents and the tests provided here, Table 8 lists the core hypotheses and the date of their generation along with a mapping to the current measures (number 1 - 16).

Second, we altered the test on 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* after accounting for the effect of the share of funds missing on complaints (technically we looked for the marginal effect of *Tuungane*, controlling for funds missing). Here we seek to examine how *Tuungane* affects the propensity to complain in light of funds missing (technically we looked for the interactive effect of *Tuungane* and funds missing). This approach we feel is more faithful to the hypotheses being examined, however we note that significant results were found under the revised approach but not under the original approach.

Third, the index on health information flows was changed to focus only on items that were provided to peers (excluding items provided uniquely to chiefs). This was to reflect the intention of the original measure but produces no substantive effect on results.

Fourth, for a number of complex tables we added summary analyses, generally mean effects analysis, as described in Section 3.3. These make for easier interpretation of the multiple results described in given tables.

Note finally that the registration document and mock report covered the main hypotheses and tests; the tests provided in Section 5 were not elaborated in the mock report.

Table 8: Hypotheses Published Prior to Data Collection

#	Category	Hypothesis	Date	Measures (Table 2)
G1	Participation	Individuals in *Tuungane* communities will participate more in collective decision making.	2007	1,2,3,5
H11/G3	Participation	Individuals in *Tuungane* communities are more likely to believe that local leaders should be elected rather than selected through an alternative mechanism.	2007	4
H10/G2	Accountability	Individuals in *Tuungane* communities will report an increased willingness to hold traditional and political leaders accountable.	2007	6,7,8
HR1	Efficiency	Projects will be implemented more efficiently in *Tuungane* areas.	2010	9, 10
H9/G	Transparency	Individuals in *Tuungane* communities will report greater knowledge about local decision-making processes and outcomes.	2007	11, 12
HR2	Capture	Benefits will be more broadly distributed in *Tuungane* communities	2010	12-16

Notes: Hypotheses H9-H11 were generated in 2007 prior to the intervention. These hypotheses were operationalized in 2010, before data collection. The operationalizations are found in Humphreys, Sanchez de la Sierra and Van der Windt (2011).Other hypotheses were added in that phase. Additional 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.

These changes are not small but they are done on the principle that it is better to use preanalysis plans to allow for implementation of better analyses transparently than to use them to constrain analysis. Finally we emphasize that these deviations do not substantively alter the results and that all results are available as originally specified at Humphreys, Sanchez de la Sierra and Van der Windt (2012) .