Stanford | King Center on Global Development

Long Run Effects of Aid: Forecasts and Evidence from Sierra Leone

Katherine Casey

Rachel Glennerster

Edward Miguel

Maarten J. Voors

Jul, 2021

Working Paper No. 1089

NBER WORKING PAPER SERIES

LONG RUN EFFECTS OF AID: FORECASTS AND EVIDENCE FROM SIERRA LEONE

Katherine Casey
Rachel Glennerster
Edward Miguel
Maarten J. Voors

Working Paper 29079
http://www.nber.org/papers/w29079

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2021

We thank the Decentralization Secretariat, the GoBifo Project, Local Councillors in Bombali and Bonthe districts, and a panel of experts for their collaboration. We thank Samuel Asher, Angelica Eguiguren, Erin Iyigun, Andres F. Rodriguez, Mirella Schrijver, Eleanor Wiseman and the Innovations for Poverty Action team in Freetown for excellent research assistance and fieldwork. We thank Macartan Humphreys, Stefano Della Vigna, Eva Vivalt, and numerous seminar participants and the 2018 BITSS Forecasting Conference for valuable comments. We gratefully acknowledge financial support from the UK Economic and Social Research Council, the Governance Initiative at J-PAL, NWO 451-14-001 and the Stanford Institute for Innovation in Developing Economies. Human subjects approval was obtained from the Sierra Leone Ethics and Scientific Review Committee, Stanford University (#38846), MIT COUHES (#1612798296) and Wageningen University. This RCT was pre-registered in the American Economic Association Registry for randomized control trials under trial number 1784. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Long Run Effects of Aid: Forecasts and Evidence from Sierra Leone
Katherine Casey, Rachel Glennerster, Edward Miguel, and Maarten J. Voors
NBER Working Paper No. 29079
July 2021
JEL No. H41,I25

ABSTRACT

We evaluate the long-run effects of a decentralized approach to economic development, called community driven development (CDD), a prominent strategy for delivering foreign aid. Notably we revisit a randomized CDD program in Sierra Leone 11 years after launch. We estimate large persistent gains in local public goods and market activity, and modest positive effects on institutions. There is suggestive evidence that CDD slightly improved communities' response to the 2014 Ebola epidemic. We compare estimates to the forecasts of experts from Sierra Leone and abroad, working in policy and academia, and find that local policymakers are overly optimistic about CDD's effectiveness.

Katherine Casey
Stanford University
Graduate School of Business
655 Knight Way
Stanford, CA 94305
and NBER
kecasey@stanford.edu

Edward Miguel
Department of Economics
University of California, Berkeley
530 Evans Hall #3880
Berkeley, CA 94720
and NBER
emiguel@econ.berkeley.edu

Rachel Glennerster
Foreign, Commonwealth and
Development Office
22 Whitehall
London, SW1A 2EG
United Kingdom
rachel.glennerster@fcdo.gov.uk

Maarten J. Voors
Development Economics Group
Wageningen University
Hollandseweg 1, 6706 KN, Wageningen
The Netherlands
maarten.voors@wur.nl

1. Introduction

Since the 1990s, community driven development (CDD) has emerged as a dominant approach to distributing foreign aid to poor and vulnerable communities. At its core, CDD devolves control over the selection, implementation and management of local public goods to communities (White 1999, Mansuri and Rao 2013). This highly decentralized and participatory approach has two main goals: to bolster local public infrastructure and associated economic activity through the provision of block grants; and to democratize community decision-making via social facilitation focused on the inclusion of marginalized groups. As a leading donor, the World Bank alone spent $85 billion over the first two decades of CDD programming (Mansuri and Rao 2013), and currently maintains $19.2 billion in active investments across 78 countries (Wong and Guggenheim 2018).

Meta-analysis of recent field experiments suggests that CDD effectively delivers local infrastructure, accompanied by little discernable impact on institutional outcomes, at least in the short run (Casey 2018). There is almost no data on how CDD performs over the longer term. This is an important lacunae to fill in light of the often elusive nature of aid sustainability (Kremer and Miguel 2007), and the open question of whether external reforms to strengthen institutions can indeed succeed when afforded a sufficiently long time horizon. CDD offers an instructive application for these questions, given its policy prominence and commensurate resource allocation, as well as the fact that early programs are now "aging" into a stage where it is possible to assess longer run effects (Bouguen et al 2018).

This study makes three primary contributions. First, it experimentally evaluates the impacts of a high-profile CDD program in Sierra Leone more than a decade after implementation began, using an array of measures to capture public goods provision, economic activity, social capital and local institutions. Second, it assesses whether the public infrastructure and collective

organizing support provided by the CDD program enabled communities to better prepare for and more effectively respond to the hardships brought on by the 2014 Ebola public health crisis. And third, it compares program effects observed on the ground with the prior beliefs and forecasts of a large number of experts located in Sierra Leone and abroad, and working in both policy and academia.

Our analysis centers on the "GoBifo"[1] CDD program, which was implemented by the Government of Sierra Leone's Decentralization Secretariat with support from the World Bank. A first intense phase of the program ran from 2005 to 2009, where treatment communities each received roughly $5,000 in block grants and six months of dedicated social facilitation. Participating communities established village development committees (VDC), mandated to include representatives of marginalized groups, which were trained and encouraged to make the selection and implementation of local projects in a democratic manner. Program staff closely monitored community observance of these participation and inclusion rules, and both their administrative records and our survey data document widespread adherence. VDC members then had the opportunity to learn-by-doing in managing a series of small-scale public projects funded by the grants, and liaised regularly with members of local elected government. A second less intense phase of the program commenced in 2010, which provided additional grants to a subset of treatment communities and continued some lighter touch engagement with project staff.

This is an informative context to study the long run effects of CDD (our first contribution). To start, the treatment was relatively intense, well-implemented and impactful in the short run. In earlier work, we found substantial positive impacts on local public goods and economic activity, and stronger links between the community and local government, over the first 4 years of program

[1] "GoBifo" means "move forward" in the local Krio language.

activity (Casey, Glennerster and Miguel 2012). Given the high rates at which aid-funded infrastructure has been found to fall into disrepair in similar contexts (Miguel and Gugerty 2005), it is useful to assess whether public infrastructure provided under CDD fares any better, particularly as it is constructed at relatively low cost (Wong 2012).

Our earlier work also found precisely estimated null results of CDD on a broad range of measures capturing institutional change, a finding that has since been challenged on both theoretical and econometric grounds, which thus provides a further motivation for a longer term follow-up. Conceptually, some critics argue that the initial evaluation timeline may have been too short to capture impacts on slowly evolving institutions, especially if institutional change follows a non-linear trajectory (Woolcock 2013). Statistically, Anderson and Magruder (2017) reanalyze the earlier data using more flexible, and thus higher powered, econometric methods, and find support for positive short-run effects of CDD on participation in local governance, which is one of several institutional dimensions examined. Partially in response to these perspectives, we returned (in 2016) to all 236 originally sampled communities, seven years after the short-run data collection (in 2009) and eleven years after program launch (in 2005), in order to assess long-run changes in institutions, and evaluate the persistence of CDD investments in local public goods.

The second contribution of this paper is to investigate the transferability of efforts to build infrastructure and promote collective action in coping with unanticipated shocks. The 2014 outbreak of the Ebola Virus Disease in West Africa is the largest ever recorded. The crisis resulted in over 4,000 deaths in Sierra Leone alone (of roughly 11,000 in total in the broader region). Some of the actions the government asked communities to take to prepare for and respond to cases—such as create community by-laws, report suspected cases, and disseminate prevention information—could be facilitated by local institutional capacity of the kind GoBifo aimed to build,

which our experimental design enables us to evaluate.

Third and finally, we elicited the prior beliefs of experts about the prospects for long-run change, and compare their predictions to our empirical estimates. We collected this data in 2016-17 which, to our knowledge, is among the first such elicitations for a field experiment. This enables us to assess the accuracy and variability of well-informed forecasters in this context. Specifically, we collected priors regarding the long run effects of CDD on both institutional and infrastructural outcomes from 126 experts familiar with CDD, a group that includes practitioners in Sierra Leone and multi-lateral institutions, like the World Bank, as well as research faculty in economics and political science, and their graduate students. This exercise adds a few data points to broader efforts to systematically document prior beliefs and compare them to outcomes obtained in lab and field settings (see Della Vigna and Pope 2018, Vivalt and Coville 2020, Vivalt et al. 2021).[2]

2. Material and methods

2.1 Intervention and Research Timeline

This experiment tracks 236 communities in Sierra Leone over an 11 year period (see Figure 1). The communities are located in two districts, Bombali and Bonthe, which were selected with an eye toward balancing regional diversity, political affiliation and ethnic composition, while simultaneously targeting poor rural areas that had previously received little aid. Half of these communities were randomly assigned to participate in the GoBifo CDD program and the remaining half were assigned to the control group that received no assistance. Baseline data was collected in 2005 before program activity commenced.

Program implementation began with facilitators helping treatment communities assemble

[2] A platform has been established to collect these forecasts systematically: see Della Vigna, Pope and Vivalt (2019) and https://socialscienceprediction.org/.

a village development committee (VDC), which was required to include both women and young men (both considered marginalized groups), and then training VDC members in methods to select, plan, implement and monitor local development projects in an inclusive and democratic way. The first intense phase of GoBifo (2005 to 2009) disbursed roughly $5,000 per treatment community, or approximately $100 per household, for use in constructing small-scale public goods (like latrines, community centers, and cement floors for drying agriculture produce) or enterprise support (like training and start-up capital for carpentry and garment dying). During weekly visits, GoBifo staff conducted trainings, facilitated meetings, and tracked participation in program activities. Accumulated over the course of the first few years of the program, these visits and trainings translated into six months of dedicated in-person support per community. The motivating idea for these investments in local organizational capacity is that they could permanently lower the fixed cost of collective action—which could be applied to future community decisions and development activities—and thereby place communities on a stronger development trajectory that outlasts the direct financing stage.

To capture the short- to medium-run impacts of the program, the research team collected data in 2009. Data collection was organized under 12 hypotheses about how CDD could alter community outcomes: 3 of these hypotheses concern the "hardware" of development, like public goods provision and economic activity; and the remaining 9 capture measures of institutional "software," like social capital, inclusion and participation (see Table A2 for a detailed list). These hypotheses were developed in partnership with the CDD practitioner team in 2005. Casey, Glennerster and Miguel (2012) analyze the 2009 data and find strong positive impacts for the hardware family of outcomes, and a series of precisely estimated null results for the software family. This empirical pattern broadly resonates with other short-run experimental studies of CDD

programs in Afghanistan (Beath, Christia and Enikolopov 2013), the Democratic Republic of Congo (Humphreys, Sanchez de la Sierra and van der Windt 2019) and Liberia (Fearon, Humphreys and Weinstein 2015).

After this first assessment, a second less intensive phase of GoBifo began in 2010. The program disbursed additional grants to 60 of the 118 treatment communities, amounting to $1,300 per community to support youth empowerment activities ("youth" is defined by the government as individuals under 35 years of age).[3] Once again, no activities were implemented in the control communities. Facilitation staff in both district headquarters (as well as management staff in the capital) were employed full time throughout this second period, and remained on government payroll at least through the long-run data collection, in 2016-18. They continued some project facilitation activities in treatment villages, although we lack reliable data on the frequency of these interactions, and our impression is that the level of program support for treatment villages was minimal after 2012.

In 2016, field enumeration teams returned to the original sample of 236 villages in order to collect long-run data, covering both the original 12 research hypotheses as well as a new hypothesis about community responses to the 2014 Ebola epidemic. Analysis in this paper thus evaluates the persistence of the initial financial and organizational investments made under the first intense phase of Gobifo, plus any additional effects of the subsequent treatment "dose" delivered in the second phase. Total project costs for the first phase (2005-2009) are approximately $2 million, and for the second, less active phase (2010-2018) nearly $3 million, given the continuation of project staffing, transport and overhead for several years. From a broader policy perspective, we evaluate a $5 million investment in CDD that was at least nominally operational

[3] This subset of 60 of the treatment communities was not randomly selected.

for more than a decade.

2.2 Long-run Data Collection

The 2016 long-run data collection aimed to replicate as closely as possible the infrastructure and institutional measures collected in 2009, as well as extend consideration to new measures capturing community responses to the Ebola crisis. To do so, field teams conducted focus group discussions with local leaders, and physically inspected a suite of community amenities and observable indicators of market activity. Note that while the 2009 data collection included both household- and community-level surveys, budgetary constraints limited the 2016 collection to community-level outcomes only. Where possible, we include community-level analogues of omitted household-level indicators, however the set of indicators collected in 2016 remains a subset of that collected in 2009. We pre-registered all outcomes and analysis in the AEA registry (see https://www.socialscienceregistry.org/trials/1784 and pre-analysis plan in Appendix C).

To bolster our ability to accurately capture the somewhat elusive concept of local institutions, we supplement survey indicators with observed communal behavior. This exercise aims to loosely replicate the structured community activities (or "SCAs") that we developed in 2009 and discuss in Casey, Glennerster and Miguel (2012). As background, the 2016 activities relate to a project challenge competition that the elected district councils were running at the time. This competition awarded $2,500 grants to support local public infrastructure projects, selected based on the quality of proposals submitted by communities. To publicize this opportunity, field team supervisors held a public meeting in all study communities. Supervisors explained what was required to enter the competition (namely, developing a project idea and completing a standardized but somewhat technical 3-page proposal), and asked community members to nominate five people

who had the requisite skills to lead the community through the proposal process. The enumeration teams then stood back outside the meeting and allowed communities to deliberate as they saw fit. Enumerators discretely observed the ensuing proceedings and recorded information on how the deliberation unfolded, the presence and engagement of youth and women, and the influence of local leaders on the process. These measures of observed behavior expand and deepen our analysis of local institutional performance.

2.3 Expert Prior Elicitation

To capture the expectations of experts about what the 2016 data collection would reveal, we asked knowledgeable policy makers and academics to make a series of predictions before we analyzed any of the data. The expert survey asked for forecasts in three areas: the long-run effects of CDD on (i) infrastructure and (ii) institutions, and (iii) the response of communities to the district government grants competition.

We first asked experts to make predictions about the same 12 hypotheses that we used in our earlier work. For each hypothesis, the survey instrument restates the hypothesis (e.g. "Hypothesis 1: GoBifo Project Implementation"), provides an example of indicators used to measure the hypothesis (e.g. "Examples of indicators include the presence of a village development committee and formal bank account for village project expenses"), and asks for a prediction about the long-run results using a slider bar that ranges from -0.50 to +0.50 standard deviation units (sdu's) (see instrument in Appendix A). As not all experts are familiar with this metric, the survey describes what standard deviation units are and provides rules of thumb for what constitutes small versus large effects. We randomly varied whether or not the survey prompted the expert with the medium run results about CDD (e.g. "our study found medium-run effects for this

hypothesis equal to +0.20 sdu's, which is statistically different from zero with a very high degree of confidence").

We then asked experts about the grants competition. This section of the survey provided background information on the competition and the procedures the field supervisors followed in publicizing it to communities, including the process for generating nominations for local residents who could lead the proposal process. The survey then asked for predictions about what percentage of communities would enter the competition.[4]

One strength of the research design is the broad variety of experts who participated, including those in the country being studied. Through systematic outreach we collected priors from 126 experts, including policymakers in Sierra Leone with knowledge of the GoBifo project; policy experts working for multilateral aid agencies such as the World Bank, primarily based in Organization for Economic Co-operation and Development (OECD) countries; faculty in both economics and political science who have been involved in evaluating CDD projects or related areas of development (including ourselves, the co-authors of this article); and economics students in Sierra Leone (undergraduates) and OECD countries (doctoral students). The variety of experts surveyed allows us to test some commonly held views about whether policymakers and academics have similar levels of optimism and/or bias regarding intervention impacts, and potential regional differences in perspective between international experts versus those located in the host country.

2.4 Empirical Strategy

To assess the long-run impacts of CDD we estimate the following model:

[4] This survey section also references a new experiment that we overlaid on top of this long-run CDD sampling frame, which is analyzed in Casey et al (forthcoming).

$$Y_c = \beta_0 + \beta_1 CDD_c + W'_c\Psi + X'_c\Gamma + \varepsilon_c \quad (1)$$

where outcome Y (e.g. presence of a public good, institutional outcome, or Ebola response measure), is measured for each community c; *CDD* is an indicator for participation in the long-run GoBifo program; W_c is a vector of stratification fixed effects for geographic wards; X_c are balancing variables used in the original 2005 randomization (community size and distance to nearest road); and ε_c is an idiosyncratic error term. We further test for heterogeneous treatment effects along the same eight community-level dimensions we used (and measured) in the short-run analysis (namely, total households, war exposure, average schooling, distance to road, historical domestic slavery, district, ethnic fractionalization and chiefly authority, see Appendix Table A3).[5]

Throughout the analysis, we adjust for the fact that we conduct multiple tests on the same dataset by implementing false discovery rate (FDR) corrections (see Benjamini, Krieger and Yekutieli [2006] and Anderson [2008]). These adjustments run across the two outcome families, or across hypotheses within a given family, as relevant. We also report the "naïve" or "per comparison" *p*-value for those interested in a particular hypothesis on its own.

We also test directly for decay in the estimates from the short- to long-run using the following model:

$$Y_c^L - Y_c^S = \gamma_0 + \gamma_1 T_c + X'_c\Lambda + W'_c\Theta + \mu_c \quad (2)$$

where the dependent variable is the difference in mean effects indices measured in the long-run survey, Y_c^L, and short run, Y_c^S. The coefficient of interest is γ_1, where $\gamma_1 < 0$ suggests that the treatment effect has dissipated over time for that outcome. Note that the set of outcomes varies between the 2009 and 2016 data collection rounds, so each index incorporates the relevant

[5] Consistent with Casey, Glennerster and Miguel (2012), we find little evidence for heterogeneous effects save for smaller impacts in one of the two study districts, namely Bombali district.

outcomes for that particular survey round (see Appendix Table A1 for estimates limited to the exact panel outcomes).

3. Results

3.1 Long-run CDD Effects

Infrastructural Hardware Outcomes

We find evidence for positive, highly significant impacts of the CDD program on measures of development hardware over the long-run. For the overall "family" of infrastructure outcomes, Table 1, Panel A reports a long-run treatment effect of 0.204 standard deviation units, which is sizeable in magnitude and statistically significant at the 99% confidence level. Estimates do not change substantively when we limit the set of outcomes to those that form an exact panel (which includes 28 of the original 39 outcomes from 2009): the 2016 treatment effect estimate is 0.208 standard deviation units (standard error 0.041) in Appendix Table A1.

This positive effect reflects gains across the three component hypotheses, which concern project implementation (e.g., does the community have a VDC?), the stock and quality of local public infrastructure (e.g., does the community have a functional water well?), and economic activity (e.g., how many goods are for sale in the community?). For each component hypothesis, the CDD treatment effect estimate is positive and large in magnitude, ranging from 0.228 to 0.253 standard deviation units. The estimates are highly statistically significant (in column 1), even after accounting for the fact that we are testing multiple hypotheses on the same dataset (in column 3). Considering decay over time, the family-level long-run effect of 0.204 standard deviation units is two thirds the size of the effect estimated in the short-run, which was 0.298 standard deviation units (in column 4). This suggests a considerable degree of persistence, even years after most

direct financial support ceased. The estimated decay, where one third of the original effect has dissipated, is statistically distinct from zero (column 5).

Looking at a few of the specific measures within each hypothesis provides a clearer sense of the magnitude for these effects. For the first hypothesis about project implementation, consider the specific measure of whether or not the community has a village development committee (VDC). In the short-run data, treatment communities were more likely to have a VDC by 40 percentage points (on a base rate of 45.8 percent in control communities, see Casey, Glennerster and Miguel 2012). In the long-run, this treatment advantage has reduced to a 17 percentage point difference (see Appendix Table A5 for treatment effect estimates for all individual outcome measures). The prevalence of a VDC in control communities has remained roughly constant over time (at 43.2 percent in 2016). These VDC estimates illustrate the broader trend for this hypothesis overall, which had both the largest estimated short-run effect (of 0.703 standard deviation units in column 4) and exhibited the strongest decay over time (of -0.449 standard deviation units in column 5) among the hypotheses in this family.

By contrast, there is no statistically detectable change from the short- to long-run for the second hypothesis about impacts of the program on the stock and quality of local public goods. At the level of individual outcomes, this effect captures enduring improvements in the availability of functional agricultural drying floors, traditional birth attendant huts, and court "barries" (or public buildings for dispute resolution), among others (see Appendix Table A5). For the third hypothesis, measures of economic welfare suggest that one third of the initial gains dissipated over time. The remaining benefits reflect persistent increases in local market activity, including enumerator observation of petty traders active in the community on the day of the 2016 field visit.

In our view, these results showing persistent gains in the "hardware" family of development outcomes are impressive, and particularly so given the challenges of working in a post-conflict environment. While our experimental design does not allow us to directly compare infrastructure provision under CDD versus other delivery mechanisms, there are some useful benchmarks in the literature. Miguel and Gugerty (2005), for example, find that nearly half of borehole water wells built by a European bilateral aid donor in Kenya in the 1980's were no longer functional within a decade of construction. Our estimated loss is only one third for CDD investments over a comparable time frame. The comparatively strong CDD performance is particularly encouraging given that CDD projects tend to be implemented at lower cost than other government service delivery mechanisms (Wong 2012), raising the question of whether they were done to a lower standard. While we cannot parse mechanisms underlying the CDD effect, these relatively favorable results are at least consistent with CDD advocates' claims about the value of local participation in aligning investments with demand and thereby bolstering utilization and maintenance over time (Dongier et al. 2002).

The Sierra Leone results provide evidence for stronger positive effects when compared to the one other longer-run CDD experiment that we are aware of in the literature, namely Mvukiyehe and van der Windt (2020) in the Democratic Republic of Congo (DRC). While they find some positive effects for the persistence of physical infrastructure, they estimate null results for long-run impacts on service delivery, economic welfare, social inclusion and local institutions. Our study is distinct from theirs in that it operates over a longer time horizon (returning 11 versus 8 years after project launch) and follows up on stronger short-run results (see Humphreys, Sánchez de la Sierra and Van der Windt 2019 on null results for the DRC program), which provides a more relevant setting for investigating the persistence of effects.

Institutional Software Outcomes

Analysis of the 2016 data yields small positive estimates for the long-run effects of CDD on local institutions. Combining all 61 individual outcomes grouped under this family into an equally weighted index yields a positive, precisely estimated, but small in magnitude treatment effect of 0.062 standard deviation units (standard error 0.024) in Table 1, panel B. When we break these estimates into the nine distinct hypotheses about how CDD might alter institutions, three hypothesis-level estimates are positive—namely for collective action, trust, and groups and networks—and at least marginally significant on a per-comparison basis (in column 2). Yet none of the hypothesis-level estimates remain significant after adjusting for multiple inference (column 3). One way to interpret this pattern of results is that if we conceive of all outcomes measuring a latent variable associated with institutional quality, CDD had a small positive impact, but the effect is not large enough to detect effects along any of the nine underlying channels.

Somewhat resonant with this interpretation, we do not find much evidence for detectable differences in behaviors observed relating to the project challenge application. Combining all 13 SCA measures, the overall treatment effect is -0.001 with standard error 0.046 (see Appendix Table A4). On a per comparison basis, only one individual indicator registers a statistically significant effect, which is a large positive effect on the time that the community took to generate its list of 5 nominees. In a companion paper, we analyze a broader array of outcomes related to the grants competition and find weak evidence for CDD effects on intermediate measures—like the village chief's willingness to delegate proposal authority to one of the community nominees—but null results for the ultimate outcomes of interest, which is winning one of the actual grants (see Casey et al., forthcoming).

How does the small positive long-run effect on institutions compare to what was measured in the short-run? While the 2016 point estimate is more than twice as large in magnitude as the null result for 2009 (0.065 versus 0.028 standard deviation units), the estimated decay over time is not statistically distinct from zero (in column 5). Yet recall that these two estimates operate over different subsets of indicators, as the long-run data collection does not include household surveys.[6] If we limit consideration to outcomes that were collected in identical fashion across the two survey rounds, the overall CDD treatment effect remains similar for 2016 (at 0.064, standard error 0.027, in Appendix Table A1). The 2009 effect, however, is somewhat larger and becomes statistically significant (at 0.086 standard deviation units, standard error 0.030). This increase in the 2009 effect could reflect differences in reporting between households and community leaders (although it is unclear to us *ex ante* which group is more or less susceptible to social desirability bias), or could be due to sampling variation created by focusing on a subset of outcomes.

3.2 CDD Impacts during a Public Health Crisis

The 2016 data collection affords a rare opportunity to measure community responses to the Ebola public health crisis of 2014 and assess whether the positive long-run effects of CDD on local infrastructure and institutions better equipped communities to cope with the associated hardship. Analysis in this section covers a variety of related outcomes, such as the creation of an Ebola task force and knowledge about the epidemic (on symptoms, transmission and control).

The estimated treatment effect for CDD on the index of the 13 combined Ebola knowledge

[6] Compared to the infrastructure family, which is based primarily on enumerator assessments of physical goods in both rounds, the lack of household data matters more here. Specifically, the 2009 round paired all community-level indicators (e.g. a count of how many people are observed at a particular community meeting) with reports from representative households (e.g. did any member of this household attend this particular meeting?), so excluding the household reports from the 2016 round cuts the number of institutional measures by roughly one half.

items and response actions is small in magnitude and not statistically distinguishable from zero (0.042 standard deviation units, with a standard error of 0.036 in Table 2). For Bombali district, which was harder hit by Ebola, the effect is also null (-0.001, 0.053, N = 156 communities, Appendix Table A6); while for Bonthe district, it is positive and significant (0.109, 0.053, N = 80). For individual outcomes, while we find no change in knowledge of Ebola, we do see some evidence that communities had taken more action. In Table 2, communities were 4% more likely to have established Ebola related by-laws (significant at 95 percent confidence) and 8% more likely to have established an Ebola task force (not significant). In a combined index of all response actions, we find a positive and marginally significant effect of 0.090 standard deviation units (standard error 0.053) in the full sample. Taken together, this provides suggestive evidence that the CDD program may have generated some benefits for villages during the Ebola crisis, although the effect magnitudes are modest.

The positive result on Ebola response actions aligns with effects found for a different, contemporaneous community program in Sierra Leone. Christensen et al (2021) analyze the impact of a randomized accountability program that facilitated community monitoring of local health clinics.[7] The program was implemented in Sierra Leone before the Ebola outbreak. Just prior to the crisis, they find that the program interventions built confidence in health workers and improved the perceived quality of care. During the crisis, this led to more reporting of Ebola cases and lower mortality from the disease. While our results are weaker, they add support for the possibility that community mobilization may be an effective strategy to generate collective action under crisis conditions.

[7] The program consisted of an additional treatment arm that provided non-financial recognition for clinic staff. Results across both arms were similar, though generally stronger for the community monitoring intervention.

3.3 Expert Forecasts

Are these results on the long-run effects of CDD in line with what informed individuals would have expected? Figure 2 presents expert predictions, collected before any data analysis, to help answer this question. The panels display forecasts for three distinct areas: CDD effects on infrastructure (Panel A), CDD effects on institutions (Panel B), and community entry into the grants competition (Panel C). For each type of expert (e.g. policymaker or academic faculty), the hollow circles portray individual expert predictions, the solid circle denotes the mean prediction for the group, and the whisker plot portrays the 95% confidence interval. We compare these forecasts to the realized effect size documented in the 2016 data, which is presented by a solid horizontal line, with dashed lines demarcating the accompanying 95% confidence interval.

Starting with long-run impacts on CDD-funded infrastructure investments, Panel A reveals that expert predictions were highly variable yet largely accurate on average. Pooled together, the experts predicted a long-run treatment effect of 0.218 standard deviation units (standard error 0.126), which is statistically indistinguishable from the estimated effect (of 0.204). A notable feature of these forecasts is the wide dispersion—ranging from zero to 0.5 standard deviation units—which is evident both within and across the different types of expert. It further appears that policymakers in Sierra Leone were relatively more optimistic about persistent infrastructure gains and faculty more pessimistic. The predictions of economics students track those of policymakers in their respective regions.

For institutions, the most striking feature of Panel B is how relatively optimistic experts in Sierra Leone were about the scope for long-run impacts. Policymakers and students alike in Sierra Leone predicted average effects in the range of 0.25 standard deviation units, which turned out to be a substantial overestimate compared to the realized effect size (of 0.066). Policymakers and

students in the OECD on average were roughly on target. While we cannot reject that economics and political science faculty were correct on average, they were more pessimistic: a substantial number of them (11 out of 23) predicted precisely zero long-run effects, which falls outside the 95% confidence interval of the observed point estimate.[8] If we pool all expert predictions together, the long-run forecast for institutional change significantly exceeds what was estimated in the short-run (0.095 predicted by experts, compared to 0.028 units in Casey, Glennerster and Miguel 2012). This difference remains statistically distinct from zero even when limited to the subgroup of experts who were randomly chosen to be primed with additional information on the short-run results (results not shown).

The substantial *ex ante* disagreement among seemingly well-informed experts about CDD's long-run institutional impacts makes the 2016 data collection an interesting empirical exercise, and particularly so in light of the accumulation of shorter-run null results for institutional outcomes from several studies (see Wong 2012, King and Samii 2014, White et al. 2017, and Casey 2018 for cross-country reviews). Moreover, the divergence between policymakers in Sierra Leone and academics lends some credence to concerns about optimism bias among policymakers and gripes (from policymakers) about hard-to-please academics, although note the substantial variation in priors among both types of expert. This potential disconnect does not appear to be as severe for policymakers based in the OECD countries, suggesting that the feedback loop between academic results and policy perceptions may be working relatively well for policymakers who are more proximate to rich country scholars, perhaps due to more frequent interactions at conferences and policy fora.

[8] The co-authors of this paper, whose forecasts are excluded from Figure 2, predicted more pessimistic long-run outcomes with an average of 0.147 (standard deviation 0.144) for hardware outcomes and 0.008 (standard deviation 0.017) software outcomes.

By contrast, all expert opinion diverged substantially from observed outcomes regarding entry into the infrastructure grants competition. As a group, the experts predicted a baseline take up rate of 42 percent, which reflects the sentiment of one expert who cautioned that "it is very likely that $2,500 is just too small an amount to get enough communities to bother with applying." In practice, we found a take up rate of 98%, which surprised all experts and far exceeded any prediction in the sample (in Panel C). Appendix Table A7 shows that experts on average expected CDD treatment communities to take up the grants opportunity at slightly higher rates than controls (by 7 percentage points), a difference that we do not observe in practice.

4. Discussion and Conclusion

Community driven development commands a substantial share of foreign aid allocations. Its short-run effects have been fairly extensively studied by randomized controlled trials in several different countries. This study broadens the evidence base by: (i) extending the time horizon to capture longer run effects than any existing study in the literature (to our knowledge), (ii) evaluating impacts during a subsequent public health crisis, and (iii) comparing expert forecasts to observed impacts.

First, following up with communities more than a decade after baseline data collection, we document strong persistence of CDD impacts on local public infrastructure, commensurate with two-thirds of the short-run gains (measured seven years prior). We find modest positive long-run effects on local institutions, which runs contrary to our own prior beliefs, although it is likely that these effects are too small (+0.062 standard deviation units on average) to be of much practical consequence.

Second, we find suggestive evidence that these positive long-run effects on infrastructure

and, to a lesser extent, institutions helped communities respond somewhat more effectively to the 2014 Ebola epidemic. While it is too early to understand the effect that GoBifo may have had on community preparedness and outcomes during the ongoing (at time of writing) COVID-19 pandemic, nor do we have the data to do so, this finding from the Ebola crisis opens the possibility, at least speculatively, that earlier CDD programming may translate into positive gains during crisis.

Finally, comparing the empirical estimates to expert forecasts, we find wide dispersion in prior beliefs, a high degree of accuracy for some types of experts on particular outcomes, accompanied by systematic underestimation for others. Taken together, the forecasts offer a few data points on the question of when and how expert predictions may be useful in research: we see (i) wide dispersion of views regarding the durability of infrastructure, (ii) disagreement across expert type for institutional change, and (iii) systematic underestimation for community entry into the grants competition. One striking pattern is the consistent optimism regarding this type of foreign aid among Sierra Leonean policymakers, in contrast to the overall pessimism among researchers. This could be problematic if their sanguine view of institutional change drives the continued popularity of CDD programming. If, by contrast, policymakers are primarily motivated by the positive infrastructural effects, this would be less of a concern.

While expert prior opinions may be useful for predicting some effects but not others, it remains unclear (to us) how to distinguish these cases *ex ante*. As more studies collect prior beliefs about the efficacy of policy interventions, a practice that is gaining some traction, the research community will be able to build a more thorough understanding of what types of impacts experts can reliably predict, and which types of experts—those with country knowledge, for instance, practitioner experience or academic training—are most accurate.

References

Anderson, M, 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedaian, Perry Preschool, and Early Training Projects", *Journal of the American Statistical Association*, 103 (484): 1481–1495.

Anderson M.L., and J. Magruder, 2017. "Split-Sample Strategies for Avoiding False Discoveries", NBER Working Paper 23544

Beath A, Christia F, Enikolopov R. 2013. "Do Elected Councils Improve Governance? Experimental Evidence on Local Institutions in Afghanistan" MIT Political Science Department Research Paper No. 2013-24.

Benjamini, Y., A. Krieger, and D. Yekutieli, 2006. "Adaptive Linear Step-Up Procedures That Control the False Discovery Rate", *Biometrika*, 93: 491–507.

Bouguen, A., Y. Huang, M. Kremer and E. Miguel, 2018. "Using RCTs to Estimate Long-Run Impacts in Development Economics", *Annual Review of Economics*, 11: 523-561.

Casey, K. R. Glennerster, E. Miguel and M. Voors, forthcoming. "Skill Versus Voice in Local Development", *Review of Economics and Statistics*.

Casey, K., 2018. "Radical Decentralization: Does Community Driven Development Work?" *Annual Review of Economics*, 10: 139-165.

Casey, K., R. Glennerster and E. Miguel, 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Pre-Analysis Plan", *Quarterly Journal of Economics*, 127(4):1755-1812.

Christensen, D., O. Dube, J. Haushofer, B. Siddiqi and M. Voors, 2021. "Building Resilient Health Systems: Experimental Evidence from Sierra Leone and The 2014 Ebola Outbreak", *Quarterly Journal of Economics*, 136(2): 1145–1198.

DellaVigna, S. and D. Pope. 2018. "What Motivates Effort? Evidence and Expert Forecasts",

The Review of Economic Studies, 85(2): 1029–1069.

DellaVigna, S., D. Pope and E. Vivalt, 2019. "Predict Science to Improve Science", *Science*, 366 (6464): 428–29

Dongier, P, Domelen JV, Ostrom E, Rizvi A, Wakeman W, et al., 2002. Chapter 9: Community-Driven Development. In *Volume 1 – Core Techniques and Cross-Cutting Issues*. World Bank, Washington D.C.

Fearon J., M. Humphreys and J. Weinstein, 2015. "How Does Development Assistance Affect Collective Action Capacity? Results from a Field Experiment in Post-Conflict Liberia" *American Political Science Review* 109(3):450-69.

Humphreys, M., R. Sánchez de la Sierra, and P. Van der Windt, 2019. "Exporting democratic practices: Evidence from a village governance intervention in Eastern Congo", *Journal of Development Economics*, 140: 279-301.

King E, Samii C. 2014. Fast-Track Institution Building in Conflict-Armed Countries? Insights from Recent Field Experiments. *World Development* 64:740-54.

Kremer, M. and E. Miguel, 2007. "The Illusion of Sustainability", *Quarterly Journal of Economics*, 122(30): 1007–1065.

Mansuri G, Rao V. 2013. Localizing Development: Does Participation Work? *Policy Research Report*. Washington, D.C.: World Bank.

Miguel, E., and M.K. Gugerty, 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya", *Journal of Public Economics* 89:2325-68.

Mvukiyehe, E., and P. van der Windt, 2020. "Assessing the Longer Term Impact of Community-Driven Development Programs Evidence from a Field Experiment in the Democratic Republic of Congo", World Bank Policy Research Working Paper 9140

Vivalt, E., and A. Coville, 2020. “How Do Policy-Makers Update Their Beliefs?” Working Paper.

Vivalt, E., A. Coville and S. KC. 2021. “Weighing the Evidence: Which Studies Count?” working paper

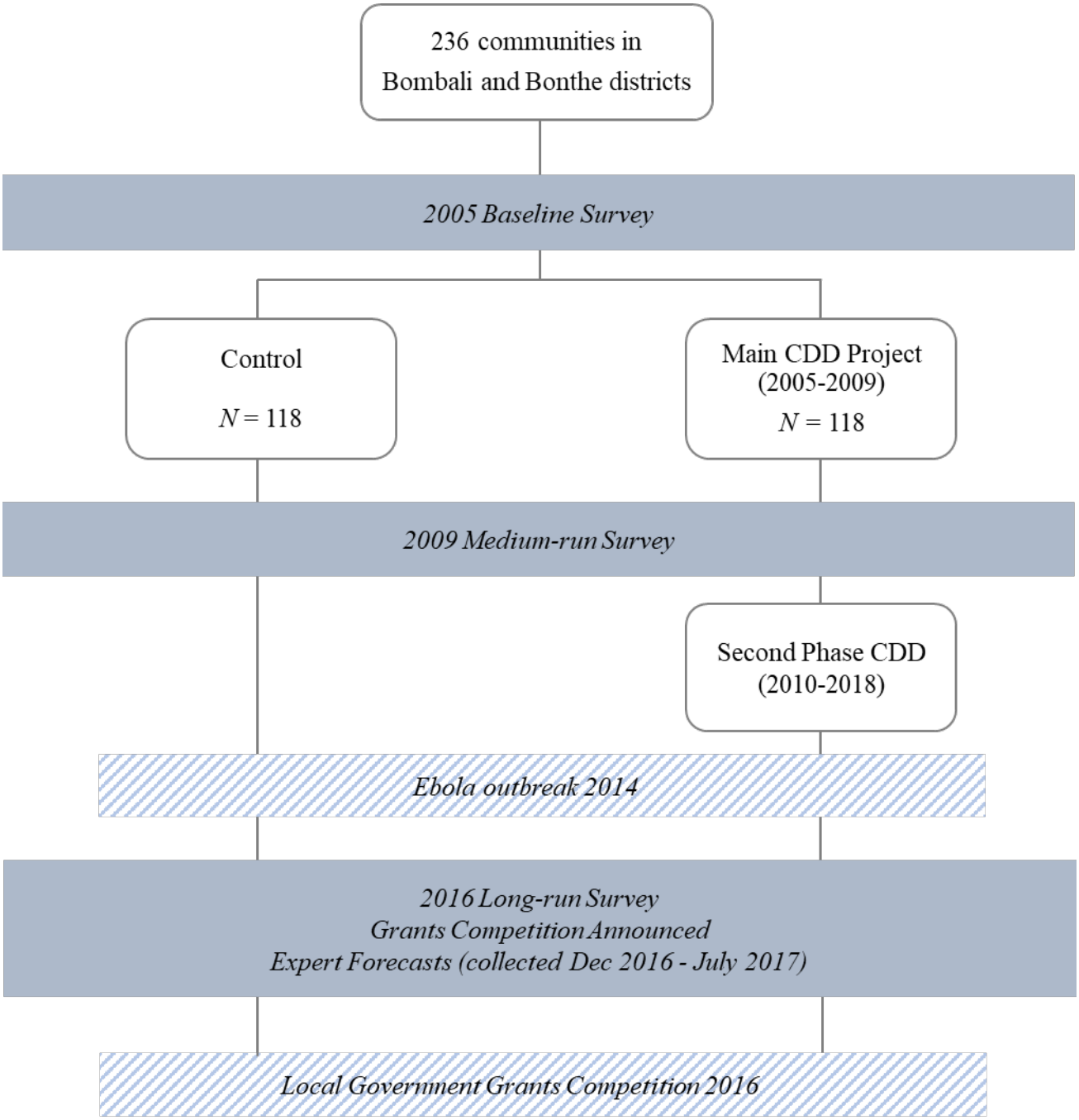
White, H. 1999. “Politicizing Development? The Role of Participation in the Activities of Aid Agencies” In: *Foreign Aid: New Perspectives* ed. Kanhaya L, Gupta. Boston: Kluwwe Academic Publishers.

Woolcock, M. 2013. “Using Case Studies to Explore the External Validity of 'Complex' Development Interventions”, *Evaluation*, 19(3): 229–248.

Wong S. 2012. “What Have Been the Impacts of World Bank Community-Driven Development Programs? CDD Impact Evaluation Review and Operational Research Implications” Washington D.C.: World Bank.

Wong, Susan; Guggenheim, Scott. 2018. “Community-Driven Development: Myths and Realities”, World Bank Policy Research Working Paper; No. 8435.

Figure 1: Research Design and Timeline



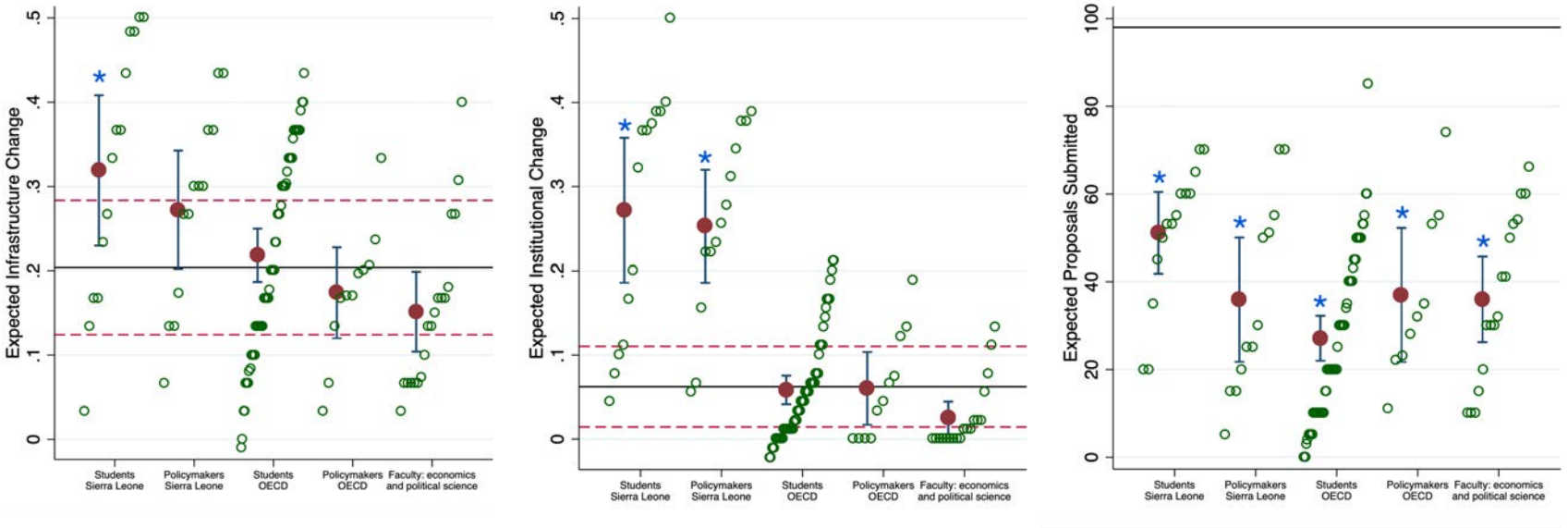
Notes: CDD treatment assignments are displayed in rounded boxes, research activity and data collection in solid gray shaded boxes, and external events and activities in dashed boxes.

Figure 2: Expert Predictions of Long-run CDD Effects and Grants Competition

Panel A: Long-run Infrastructure Change

Panel B: Long-run Institutional Change

Panel C: Entry into the Grants Competition



Notes: This figure presents expert predictions collected during December 2016 and July 2017 before any data analysis. Panels A and B present expectations for CDD treatment effects measured in standard deviation units. The realized effect size is presented with solid horizontal lines and the accompanying 95% confidence interval is demarcated by dashed horizontal lines. Panel C presents expectations about the percent of CDD control communities that would enter the grants competition. The realized point estimates are: a) 0.204 standard deviation unit CDD treatment effect for infrastructure in Panel A; i) 0.062 standard deviation unit CDD treatment effect on institutions for Panel B; and c) 98.3% percent of communities entered the grants competition for Panel C. For Panels A and B, expert predictions were closer to the realized value for the version of the survey that provided the short to medium run results for institutional change (p-value <0.01) but not statistically distinct for infrastructure (p-value=0.27). Stars above the 95 percent confidence interval denote forecasts that are significantly different from the realized effect.

Table 1: Long Run CDD Treatment Effects

	Treatment effect 2016	Naïve *p*-value	FDR *q*-value	Treatment effect 2009	Change over time
	(1)	(2)	(3)	(4)	(1) - (4)
Panel A: Infrastructure "Hardware" Family					
All outcomes (30 unique outcomes)	0.204***	<0.01	0.001	0.298***	-0.094***
	(0.040)			(0.031)	(0.036)
Project implementation	0.253***	<0.01	0.001	0.703***	-0.449***
	(0.067)			(0.055)	(0.080)
Local public goods	0.228***	<0.01	0.001	0.204***	0.024
	(0.046)			(0.039)	(0.041)
Economic welfare	0.240***	<0.01	0.001	0.376***	-0.136**
	(0.056)			(0.047)	(0.062)
Panel B: Institutions "Software" Family					
All outcomes (61 unique outcomes)	0.062**	0.011	0.124	0.028	0.034
	(0.024)			(0.020)	(0.027)
Collective action	0.098**	0.049	0.172	0.012	0.086
	(0.050)			(0.037)	(0.061)
Inclusion	0.028	0.417	0.557	0.002	0.026
	(0.034)			(0.032)	(0.045)
Local authority	-0.032	0.573	0.563	0.056	-0.088
	(0.056)			(0.037)	(0.070)
Trust	0.107*	0.065	0.172	0.042	0.064
	(0.057)			(0.046)	(0.081)
Groups and networks	0.149**	0.038	0.172	0.028	0.121
	(0.071)			(0.037)	(0.074)
Access to information	-0.036	0.591	0.563	0.038	-0.075
	(0.067)			(0.037)	(0.072)
Participation in governance	0.079	0.191	0.276	0.090***	-0.011
	(0.060)			(0.045)	(0.065)
Crime and conflict	-0.002	0.971	0.636	0.010	-0.012
	(0.063)			(0.043)	(0.074)
Political and social attitudes	0.154	0.216	0.276	0.041	0.113
	(0.124)			(0.043)	(0.126)
Observations	236				

*Notes: i) significance levels based on naive p-values and indicated by * p<0.10, ** p<0.05, ***p<0.01, ii) specifications include strata for geographic ward and two balancing variables (distance to road and community size) from the original randomization; iii) robust standard errors; iv) all estimates are for hypothesis-level equally weighted mean effects indices, expressed in standard deviation units (see Kling, Liebman and Katz 2007); v) the dependent variable in column 5 is the difference in 2009 and 2006 indices, where the set of component measures varies across survey round (see Appendix Table A3 for exact panel specification); and vi) 2009 data sourced from Casey, Glennerster and Miguel (2012).*

Table 2: CDD Treatment Effects on Ebola Knowledge Items and Response Actions

Outcome	Mean, controls	Treatment effect	Standard error	*p*-value	FDR *q*-value
Mean Effects Index (all 13 indicators)	0.000	0.045	0.040	0.256	.
Knowledge Items					
Mean Effects Index (all 9 knowledge items)	0.000	0.015	0.050	0.766	0.621
Correctly answers "No" to "Can Ebola spread through air?"	0.856	-0.005	0.040	0.896	0.999
Correctly answers "21" to "How many days can it take for the first to symptoms arise?"	0.669	0.014	0.051	0.791	0.999
Total (of 11 possible) correct answers to questions about how one can get Ebola	5.220	0.006	0.187	0.974	0.999
Knows correct Ebola hotline number	1.000	0.000	.	.	
Total (of 10 possible) correct answers regarding how to protect yourself against Ebola	4.975	-0.051	0.201	0.801	0.999
Correctly answers "No" to "Drinking salt water can help cure Ebola?"	0.958	0.030	0.019	0.114	0.999
Correctly answers "No" to "Drinking chlorine can help cure Ebola?"	1.000	-0.009	0.009	0.321	0.999
Correctly answers "No" to "Can someone spread Ebola before they show symptoms?"	0.695	0.030	0.052	0.565	0.999
Total correct answers (of 14 possible) regarding symptoms of Ebola	7.263	-0.230	0.232	0.324	0.999
Response Actions					
Mean Effects Index (all 4 response actions)	0.000	0.090*	0.053	0.091	0.223
Community had an Ebola task force during the Ebola crisis	0.661	0.077	0.052	0.145	0.278
Community created by-laws in relation to Ebola	0.907	0.042**	0.019	0.030	0.137
Communities are more likely to go to formal health facilities (nurse, clinic)	0.924	0.014	0.030	0.632	0.729
Communities are more likely to go to formal health facilities for Ebola (nurse, clinic)	0.915	0.000	0.034	0.995	0.991
Observations	236				

*Note: i) significance levels based on naive p-values and indicated by * p<0.10, ** p<0.05, ***p<0.01. ii) specifications include strata for geographic ward and two balancing variables (distance to road and community size) from the randomization; iii) robust standard errors; and iv) this table includes 13 of 15 pre-specified primary outcomes in our PAP, excluding 2 outcomes that are observed for fewer than 20 communities in the data.*

ONLINE APPENDIX: MATERIAL NOT INTENDED FOR PUBLICATION

Long Run Effects of Aid: Forecasts and Evidence from Sierra Leone

Contents:

Appendix A. Expert Prior Elicitation Details

Before collecting and analyzing the data, we first established what experts in the field *thought* we would find. To do so, we fielded a survey among different types of experts and asked them to make predictions in three main areas: i) long-run impacts of CDD on infrastructure; ii) long-run impacts on measures of institutions; and iii) community performance in the infrastructure grants competition.

Experts came from several groups: i) policymakers working for multilateral aid agencies (including the World Bank, the Department for International Development, the United Nations Development Programme and the International Rescue Committee) located mostly in OECD countries; ii) policymakers in Sierra Leone with knowledge of the GoBifo project; iii) economics graduate students in the United States (at University of California, Berkeley) and the Netherlands (at Wageningen University); iv) economics undergraduate students in Sierra Leone (at Fourah Bay College); and v) faculty in economics and political science directly involved in evaluating CDD projects (including the co-authors of this study) and other development economics researchers. This yielded 126 completed surveys in total, composed of 25 surveys from policymakers (12 in the OECD and 13 in Sierra Leone), 78 from students (17 undergraduate and 61 graduate students), and 23 from faculty. Survey response rates were quite high for all groups (e.g. 84% for faculty and 99% for graduate students) save the OECD policymakers (39% completion).

For estimates about long run CDD impacts, the survey refers to the same twelve hypotheses and comparable empirical measures that are the focus of Casey, Glennerster and Miguel (2012). For each hypothesis, the survey asks experts to predict the point estimates we would find in the long-run, in standard deviation units, and also indicate their level of certainty for each prediction (following DellaVigna and Pope 2018). As in our earlier work, we then group these hypotheses and predictions into two main families, infrastructure and institutions. There were two versions of the survey: the first provided detailed information on our medium run results and the second asked the expert to make predictions without any information provided (see instrument on page A3). We randomized which version was given to each expert, with a few exceptions (e.g. a small subset completed both versions). Expert predictions about the infrastructure grants competition focus on entry as a proxy for overall performance.

Note that a few different versions of the survey were implemented. The version we display below is the one that includes the primes regarding the shorter run results. We flag these priming

sentences below by reproducing them in *italics*. The alternative version, without primes, excludes these priming sentences but was otherwise the same. The different colors at the start of the instrument demarcate small differences in questions across pools of expert, where (i) **black** is universal (except questions 1 and 2 which were only given to academic experts, policy experts, and the co-authors of this study); (ii) blue questions were given only to students in Sierra Leone and Berkeley; and (iii) red questions were given only to Sierra Leone policymakers and Wageningen students.

Expert Survey Instrument:
Measuring the Long-Run Effects of Community Driven Development in Sierra Leone

Researchers: Katherine Casey, Rachel Glennerster, Edward Miguel, and Maarten Voors

Date: [Month, Year]

Overview: In 2012, we published the results of an impact evaluation of a community driven development (CDD) project in rural Sierra Leone, called GoBifo. That paper focused on the medium-run effects of CDD on local economic and institutional outcomes. We now plan to implement a new research project to measure the long-run effects of that project. Before we do so, we would value your input regarding what you expect these impacts to be, and have therefore prepared this brief (roughly 10 minute) survey.

Your participation is completely voluntary and you are free to leave the survey blank if you do not wish to participate. We will maintain your confidentiality by not recording any personally identifying information about you. We foresee little benefit or risk from participation, and cannot and do not guarantee or promise that you will receive any benefits from this study. If you have any questions about this research, please contact Katherine Casey at +1 (650) 725-2167. If you have any complaints, please contact the Stanford Human Subjects Institutional Review Board (IRB) at +1 (866) 680-2906.

1. **What is your job/position title?**

   _____

1. **What is your major?**

_____

1 . What best describes your professional position and experience? (CIRCLE ONE)

a. Researcher who has worked on CDD evaluations
b. Researcher who has not worked directly on CDD
c. Development practitioner who has worked on implementing CDD projects
d. Development practitioner who has not worked directly on CDD

2. **Have you heard about the project challenge competition currently running in Bombali and Bonthe?** (CIRCLE ONE)

YES / NO

2. **In what year of your program are you?**

_____

2. **Do you have any direct professional experience in Sierra Leone?** (CIRCLE ONE) YES / NO

3. **On a scale of 1 to 10, how familiar are you with <u>our 2012 study</u> of a CDD project in Sierra Leone entitled** "Reshaping Institutions: Evidence on Aid Impacts Using a Pre-analysis Plan"

(with 1 representing having never heard of it to 10 being very familiar with the results)? (CIRCLE ONE)

1 ------- 2 ------- 3 ------- 4 ------- 5 ------- 6 ------- 7 ------- 8 ------- 9 ------- 10

Never heard of it with results — *Very familiar*

4. **On a scale of 1 to 10, how familiar are you with <u>other CDD impact evaluations</u> in low income countries**
   (with 1 representing having never heard about other CDD studies to 10 being very familiar with the results of several studies)? (CIRCLE ONE)

1 ------- 2 ------- 3 ------- 4 ------- 5 ------- 6 ------- 7 ------- 8 ------- 9 ------- 10

Never heard of any — *Very familiar with several*

5. **Do you think that the World Bank <u>should continue to support</u> community driven development (CDD) programs to the extent that it currently does?** (CIRCLE ONE)
   a. The World Bank should spend more on CDD than current amount
   b. The World Bank should maintain current levels of spending
   c. The World Bank should spend less on CDD than current amount
   d. Indifferent

Standard Deviation Unit Effect

In what follows, we will ask you to predict how large the long-run treatment effects of the Sierra Leone CDD project will be. As we measure effects across groups of outcomes, standard practice is to refer to treatment effect sizes in standard deviation units (sdu's). This makes the effect sizes comparable across outcome measures. For your reference, the following table provides a rule of thumb interpretation of the real-world magnitude of standard deviation unit treatment effects of various sizes (in absolute value):

Treatment effect size in standard deviation units (sdu's), in absolute value	**Interpretation**
0.00	No impact
0.05	Very small effect
0.10	Small effect
0.20	Moderately small effect
0.30	Moderate effect
0.40	Moderately large effect
> 0.50	Large effect

MEDIUM-RUN RESULTS AND LONG-RUN FORECASTS

The CDD Project "GoBifo" (which means "move forward" in the dominant local language) in Sierra Leone was implemented from 2005 to 2009. This project provided block grants of US$5,000 (approximately US$100 per household) to communities in rural Sierra Leone. The grants could be used for the construction of local public goods, trade skills training, and small business start-up capital. GoBifo facilitators spent an average of 6 months in each of these villages promoting democratic decision-making, the participation of socially marginalized groups (such as women and youth), and transparent local budgeting practices. In addition, 60 of these villages received a follow up grant of $1,300 in 2010 for youth empowerment programs.

The project was implemented as a randomized control trial, where 118 villages participated in the GoBifo intervention and 118 served as controls that did not receive any project assistance. The original follow-up survey of medium-run treatment effects was fielded in 2009 and evaluated impacts on 12 hypotheses which we grouped into two broad sets of indicators: a family of "hardware" effects on local public goods and economic outcomes, and a family of "software" effects including institutional and social capital measures. We are now going back to the field to measure long-run effects, a full 7 years after the program ended, and would like to know your views on what you expect the long-run effects of GoBifo are likely to be.

Since there are several individual outcome measures included under each of the 12 hypotheses, we measure the average effect across all of them after normalizing measures in standard deviation units (sdu's). Below we list all 12 hypotheses tested in the study and include examples of indicators used in the survey. We also provide you with detailed results from our 2012 study of the medium-run effects of the GoBifo project.

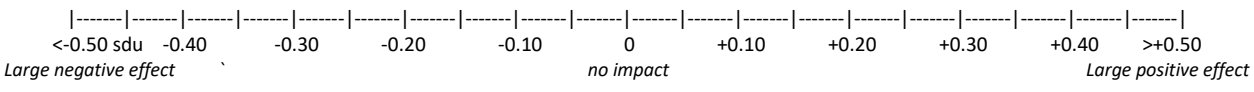
For each of 12 hypotheses below, please mark the scale with an **X** for the size of the **long-run treatment effect** of the GoBifo project that you expect we will find when we return to the field in Sierra Leone to collect data this November. We would now like to provide you more detailed results from our 2012 study of the medium-run effects of the GoBifo project, and ask you to again predict what you think the long run effects of GoBifo will be for the following hypotheses.

Hardware family of outcomes

Hypothesis 1: GoBifo Project Implementation.
Examples of indicators include the presence of a village development committee and formal bank account for village project expenses.

*Our study found medium-run effects for this hypothesis equal to **+0.70 sdu's,** which is statistically different from zero with a very high degree of confidence.*

What do you think the long run treatment effect will be?

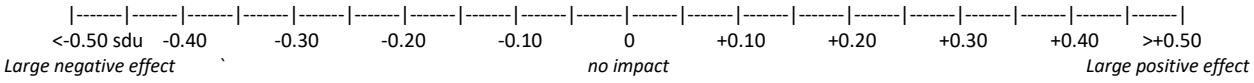


Hypothesis 2: Participation in GoBifo improves the quality of local public services infrastructure.

Examples include the presence and construction quality of latrines and drying floors.

*Our study found medium-run effects equal to **+0.20 sdu's,** which is statistically different from zero with a very high degree of confidence.*

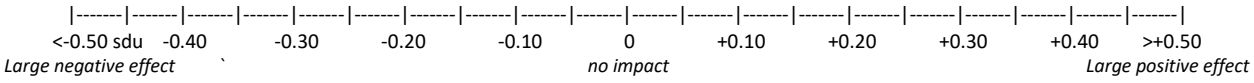
What do you think the long run treatment effect will be?



Hypothesis 3: Participation in GoBifo improves general economic welfare.
Indicators include the number of petty traders and goods on sale in the community.

*Our study found medium-run effects for this hypothesis equal to **+0.38 sdu's,** which is statistically different from zero with a very high degree of confidence.*

What do you think the long run treatment effect will be?



Software family of outcomes

Hypothesis 4: Participation in GoBifo increases collective action and contributions to local public goods.
Indicators include presence of communal farms and community-supported teachers.

*Our study found medium-run effects for this hypothesis equal to **+0.01 sdu's,** which is not statistically different than zero at traditional confidence levels.*

What do you think the long run treatment effect will be?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-0.50 sdu -0.40 -0.30 -0.20 -0.10 0 +0.10 +0.20 +0.30 +0.40 >+0.50
Large negative effect ` *no impact* *Large positive effect*

Hypothesis 5: GoBifo increases inclusion and participation in community planning and implementation, especially for poor and vulnerable groups; GoBifo norms spill over into other types of community decisions, making them more inclusive, transparent, and accountable.
Indicators include taking minutes at community meetings and reporting having fewer problems with financial misconduct.

*Our study found medium-run effects equal to **0.00 sdu's,** which is not statistically different than zero at traditional confidence levels.*

What do you think the long run treatment effect will be?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-0.50 sdu -0.40 -0.30 -0.20 -0.10 0 +0.10 +0.20 +0.30 +0.40 >+0.50
Large negative effect ` *no impact* *Large positive effect*

Hypothesis 6: GoBifo changes local systems of authority, including the roles and public perception of traditional leaders versus elected local government.
Indicators include the community choosing a village headman younger than 35 years old.

*Our study found medium-run effects equal to **+0.06 sdu's,** which is not statistically different than zero at traditional confidence levels.*

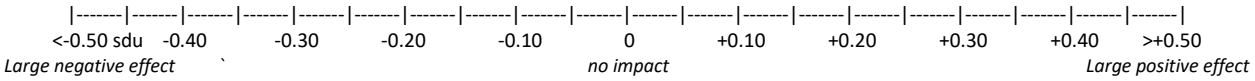
What do you think the long run treatment effect will be?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-0.50 sdu -0.40 -0.30 -0.20 -0.10 0 +0.10 +0.20 +0.30 +0.40 >+0.50
Large negative effect ` *no impact* *Large positive effect*

Hypothesis 7: Participation in GoBifo increases trust.
Indicators include the presence of cooperative trading groups that span multiple households.

*Our study found medium-run effects for this hypothesis equal to **+0.04 sdu's,** which is not statistically different than zero at traditional confidence levels.*

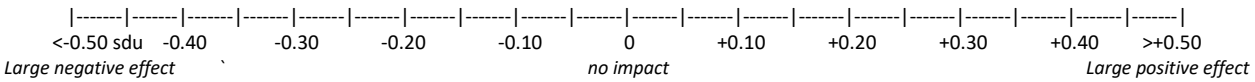
What do you think the long run treatment effect will be?



Hypothesis 8: Participation in GoBifo builds and strengthens community groups and networks.
Indicators include presence of fishing groups / cooperatives in the community.

*Our study found medium-run effects for this hypothesis equal to **+0.03 sdu's,** which is not statistically different than zero at traditional confidence levels.*

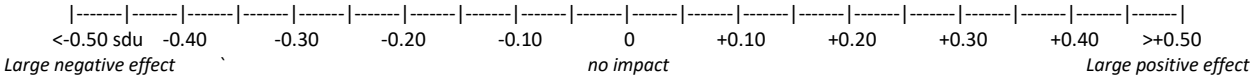
What do you think the long run treatment effect will be?



Hypothesis 9: Participation in GoBifo increases access to information about local governance.
Indicators include visits by local government officials and display of government policies or posters in the community.

*Our study found medium-run effects equal to **+0.04 sdu's,** which is not statistically different than zero at traditional confidence levels.*

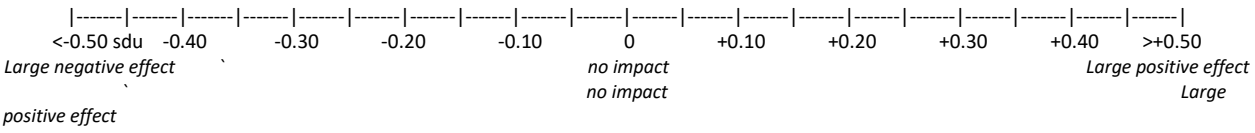
What do you think the long run treatment effect will be?



Hypothesis 10: GoBifo increases public participation in local governance.
Indicators include the involvement of local government officials in planning or overseeing community development projects.

*Our study found medium-run effects equal to **+0.09 sdu's,** which is statistically different than zero with a moderate degree of confidence.*

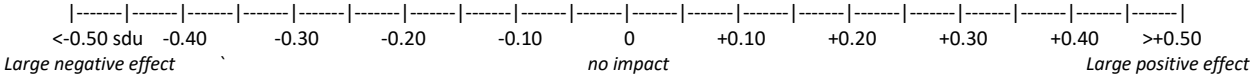
What do you think the long run treatment effect will be?



Hypothesis 11: By increasing trust, GoBifo reduces crime and conflict in the community.
Indicators include reports of theft of household items or livestock.

*Our study found medium-run effects for this hypothesis equal to **+0.01 sdu's,** which is not statistically different than zero at traditional confidence levels.*

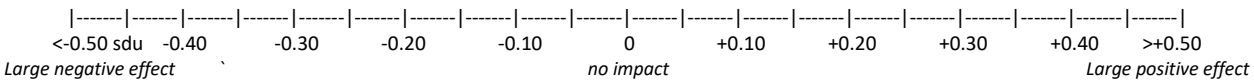
What do you think the long run treatment effect will be?



Hypothesis 12: GoBifo changes political and social attitudes, making individuals more liberal towards women, more accepting of other ethnic groups and "strangers" and less tolerant of corruption and violence.
Indicators include community choosing a woman to be the village chief.

*Our study found medium-run effects for this hypothesis equal to **+0.04 sdu's,** which is not statistically different than zero at traditional confidence levels.*

What do you think the long run treatment effect will be?



Overall expectations

You made 12 additional forecasts above about the long-run effects of GoBifo. How many of these additional forecasts do you think will fall within 10% of the true effect size (in standard deviation unit terms) that we find in the data we will begin to collect in November? _____ (out of 12)

Appendix B: Additional Specifications

Appendix Table A1: Long Run CDD Treatment Effects on Exact Panel Outcomes

	Treatment effect 2016 (1)	Naïve *p*-value (2)	FDR *q*-value (3)	Treatment effect 2009 (4)	Change over time (1) - (4)
Panel A: Infrastructure "Hardware" Family					
All outcomes in family (*N* = 29)	0.208***	<0.001	0.001	0.352**	-0.144***
	(0.041)			(0.035)	(0.037)
Project implementation	0.287***	<0.001	<0.001	0.875***	-0.588***
	(0.075)			(0.062)	(0.092)
Local public goods	0.228***	<0.001	<0.001	0.210***	0.018
	(0.046)			(0.041)	(0.041)
Economic welfare	0.240***	<0.001	<0.001	0.606***	-0.366**
	(0.056)			(0.061)	(0.062)
Panel B: Institutions "Software" Family					
All outcomes in family (*N* =56)	0.064**	0.017	0.009	0.086***	-0.021
	(0.027)			(0.030)	(0.034)
Collective action	0.104*	0.053	0.243	0.072	0.032
	(0.053)			(0.046)	(0.065)
Inclusion	0.034	0.339	0.513	0.084*	-0.050
	(0.036)			(0.049)	(0.055)
Local authority	-0.032	0.573	0.637	0.110	-0.142*
	(0.056)			(0.068)	(0.085)
Trust	0.107*	0.065	0.243	0.032	0.074
	(0.057)			(0.049)	(0.083)
Groups and networks	0.149**	0.038	0.243	0.056	0.093
	(0.071)			(0.045)	(0.080)
Access to information	-0.036	0.591	0.637	0.150**	-0.187**
	(0.067)			(0.072)	(0.092)
Participation in governance	0.079	0.191	0.35	0.256**	-0.177***
	(0.060)			(0.058)	(0.068)
Crime and conflict	-0.002	0.971	0.76	0.088	-0.090
	(0.063)			(0.062)	(0.084)
Political and social attitudes	0.154	0.216	0.35	-0.020	0.174
	(0.124)			(0.080)	(0.135)
Observations	236			236	236

*Note: i) significance levels based on naive p-values and indicated by * p<0.10, ** p<0.05, ***p<0.01; ii) specifications include strata for geographic ward and two balancing variables (distance to road and community size) from the randomization; iii) robust standard errors; iv) all estimates are for hypothesis-level mean effects indices that equally weight component measures and are expressed in standard deviation units (see Kling, Liebman and Katz 2007); v) outcomes limited to those that were collected in the exact same fashion in both 2009 and 2016 survey rounds; and vi) 2009 data sourced from Casey et al (2012).*

Appendix Table A2: Research Hypotheses

Family A: Infrastructural "Hardware"	
H1	GoBifo creates functional development committees
H2	Participation in GoBifo improves the quality of local public services infrastructure
H3	Participation in GoBifo improves general economic welfare
Family B: Institutional "Software"	
H4	Participation in GoBifo increases collective action and contributions to local public
H5	GoBifo increases inclusion and participation in community planning and implementation, especially for poor and vulnerable groups; GoBifo norms spill over into other types of community decisions, making them more inclusive, transparent and
H6	GoBifo changes local systems of authority, including the roles and public perception of traditional leaders (chiefs) versus elected local government
H7	Participation in GoBifo increases trust
H8	Participation in GoBifo builds and strengthens community groups and networks
H9	Participation in GoBifo increases access to information about local governance
H10	GoBifo increases public participation in local governance
H11	By increasing trust, GoBifo reduces crime and conflict in the community
H12	GoBifo changes political and social attitudes, making individuals more liberal towards women, more accepting of other ethnic groups and 'strangers', and less tolerant of corruption and violence
Ebola Response	
H13	Participation in GoBifo increased knowledge, collective action and investments in preventative measures during the Ebola crisis
Expert Forecasts	
H14	Estimated long run treatment effects are not the same as the average prior beliefs of surveyed experts
H15	Average prior beliefs and forecast accuracy differ across groups of experts
H16	Prior beliefs about long run effects of the GoBifo project are more optimistic (e.g. predict larger positive long run effects) amongst policy makers compared to
H17	Predictions under version 1 of the survey (that contains information on the medium run effects) are more accurate than under version 2

Notes: i) hypotheses H1 to H12 follow-up on those established for the short-run data collection (Casey, Glennerster and Miguel 2012); and ii) hypotheses H13 to H17 are new to the long-run data collection round.

Appendix Table A3: Treatment Effect Heterogeneity

	Mean Effect Index for Family A: Development Infrastructure (Hypotheses 1-3)	Mean Effect Index for Family B: Institutional and Social Change (Hypotheses 4-12)
	(1)	(2)
Treatment Indicator	0.793**	0.228*
	(0.203)	(0.122)
Treatment * Total households in the community	-0.002	-0.000
	(0.002)	(0.001)
Treatment * Index of war Exposure	-0.306	-0.005
	(0.232)	(0.143)
Treatment * Average respondent schooling	0.008	0.005
	(0.035)	(0.020)
Treatment * Distance to motorable road	0.005	0.005
	(0.017)	(0.009)
Treatment * Historical extent of domestic slavery	-0.108	-0.091
	(0.086)	(0.058)
Treatment * Bombali district	-0.457**	-0.141*
	(0.089)	(0.054)
Treatment * Ethnolinguistic fractionalization	0.122	-0.268*
	(0.239)	(0.146)
Treatment * Chiefly authority	0.141	-0.002
	(0.287)	(0.207)
Observations	236	236

*Note: i) significance levels based on naive p-values and indicated by * p<0.10, ** p<0.05, ***p<0.01.ii) robust standard errors; iii) includes fixed effects for the district council wards (the unit of stratification); iv) each specification is run on the post-program data and includes the following control variables: total households per community, distance to nearest motorable road, index of war exposure, index of history of domestic slavery, and average respondent years of school, plus all of these control variables--and the district dummy variable--interacted with the GoBifo treatment dummy; v) these mean effect estimates are limited to the full sample set of outcomes that excludes all conditional outcomes (i.e. those that depend on the state of another variable--for example, quality of infrastructure depends on the existence of the infrastructure).*

Appendix Table A4: Treatment Effect on Structured Community Activity (SCA) Outcomes

Row	Variable	Hypothesis	Mean, controls	Treatment effect	Standard error	Naïve *p*-value	N
		(1)	(2)	(3)	(4)	(5)	(6)
Index measure for all 13 SCA outcomes			0.000	-0.001	0.046	0.976	236
123	Potential managers selection deliberation done in public debate	H5	1.416	0.013	0.057	0.813	192
124	Less concentrated deliberation in manager selection	H5	2.892	0.023	0.090	0.798	231
127	Enumerator account of how democratically the group evenutally came to a decision about who the potential project managers ranging from 5 = open discussion followed by group vote to 1 = chief and/or elders decide without other input	H5	3.364	-0.002	0.094	0.982	235
130	Time of deliberation of manager selection process	H5	32.486	53.665	27.838	0.055	210
154	Enumerator account of how actively women participated in the deliberation on the selection of potential project managers compared to men, ranging from 5 = no difference between women and men to 1 = women not active at all compared to men	H5	2.799	-0.122	0.132	0.357	232
155	Enumerator account of how actively youth participated in the deliberation on the selection of potential project managers compared to non-youth (over 35 years), ranging from 5 = no difference between youth and non-youth to 1 = youth not active at all compared to non-youth	H5	3.035	0.173	0.153	0.260	229
160	Enumerator record of total public speakers durings selection of potential project managers	H5	43.429	-2.772	2.584	0.285	213
161	Enumerator record of total women public speakers during selection of potential project managers	H5	13.264	-0.570	1.176	0.628	216
162	Enumerator record of total youth (18-35 years) public speakers during selection of potential project managers	H5	6.009	-0.402	0.391	0.304	229
165	Did a vote occur during the project leader nomination discussion	H5, H6	1.929	0.023	0.032	0.464	171
201	Enumerator reports on whether "chief decided" project leader nominations	H6	0.873	-0.050	0.043	0.242	235
50	Record of total women (18+ years) in "important people" focus group list	H5	8.186	0.192	0.229	0.402	236
52	Record of total youth (18-35 years) in "important people" focus group list	H5	2.288	-0.209	0.193	0.281	236

*Note: i) Row number refers to indicator numbers as listed in the PAP, which can be found here https://www.socialscienceregistry.org/trials/1784, ii) significance levels based on naive p-values and indicated by * p<0.10, ** p<0.05, ***p<0.01. iii) specifications include strata for geographic ward and two balancing variables (distance to road and community size) from the randomization; iv) robust standard errors; and v) this table includes 13 pre-specified primary outcomes in our PAP, excluding outcomes corresponding to the grants competition (Casey et al., forthcoming).*

Appendix Table A5: Raw Results for CDD Effects on Individual Outcomes

Row	Variable	Hypothesis	Mean, controls	Treatment effect	Standard error	Naïve *p*-value	N
H1: Implementation							
1	Does this community have a bank account?	H1, H3	0.042	0.240	0.040	0.000	236
2	Average score of all test takers	H1	41.789	1.142	1.467	0.437	233
3	Does this community have a Village or Community Development Committee?	H1, H4, H10	0.432	0.173	0.057	0.003	236
4	Does this community have a village development plan (i.e. an agreed plan with specific priorities for what the community will do for its own development over the next few years)?	H1, H10	0.492	0.003	0.057	0.955	236
5	Was community visited by a Local Councillor in the past year?	H1, H9	0.263	-0.074	0.046	0.110	236
6	Was community visited by a Ward Development Committee member in past year?	H1, H9	0.102	0.019	0.035	0.579	236
H2: GoBifo improves the quality of local public services infrastructure.							
7	When was the last time this community brushed this foot path?	H2, H4	-35.224	1.123	4.707	0.812	234
8	Does the community have a court barrie and is it functional?	H2	0.102	0.218	0.040	0.000	236
9	Does the community have a community center and is it functional?	H2	0.068	0.060	0.038	0.112	236
10	Does the community have a drying floor and is it functional?	H2	0.178	0.127	0.051	0.014	236
11	Does the community have a grain store and is it functional?	H2	0.119	0.198	0.051	0.000	236
12	Does the community have a latrine and is it functional?	H2	0.076	0.029	0.036	0.413	236
13	Does the community have a market and is it functional?	H2	0.000	0.025	0.013	0.065	236
14	Does the community have a palava hut and is it functional?	H2	0.042	0.019	0.028	0.488	236
15	Does the community have a public health unit and is it functional?	H2	0.110	-0.022	0.038	0.566	236
16	Does the community have a primary school and is it functional?	H2	0.466	0.125	0.058	0.031	236
17	Does the community have any wells and are any of them functional?	H2	0.661	0.000	0.057	0.997	236
18	Do any of the local sports teams have uniforms / vests?	H2	0.153	0.003	0.046	0.946	236
19	Does the community have a football / sports field and is it functional?	H2	0.619	0.160	0.054	0.004	236
20	Does the community have a traditional birth attendant (TBA) house and is it	H2	0.025	0.124	0.032	0.000	236
21	Maintenance of bush paths. [0 "very bushy" to 1 "very clear"]	H2, H4	2.653	-0.049	0.110	0.659	236
22	Did community recently take project proposal to external funder on its own initiative?	H2, H4	0.246	0.048	0.054	0.371	236
23	Does this community have a seed bank (i.e. where people can borrow rice or groundnuts to plant and repay after harvest)?	H2	0.085	0.049	0.040	0.226	236

Appendix Table A5: Raw Results for CDD Effects on Individual Outcomes (continued)

Row	Variable	Hypothesis	Mean, controls	Treatment effect	Standard error	Naïve *p*-value	N
H3: GoBifo improves general economic welfare							
24	Supervisor assessment that community is "much better off" or "a little better off" than other communities he/she has been to in this area	H3	0.364	0.091	0.058	0.115	236
25	When was the last time an outsider trader came to this village to buy agricultural or non-agricultural goods? (date - date of interview)	H3	-12.178	3.468	4.820	0.473	236
26	[From supervisor tour of community] Have you seen anybody selling packaged goods (cigarettes, crackers, etc) in this village today from their own home (i.e. not out of a store)?	H3	0.881	-0.015	0.040	0.706	236
27	Number of goods out of 10 common items (bread, soap, garri, country cloth/garra tie-dye, eggs/chickens, sheep/goats, palm oil/nut oil, coal, carpenter for hire/shop, tailor/dressmaker, blacksmith for hire/shop) that you can buy in this community today	H3	5.619	0.403	0.247	0.105	236
28	How many people have started a new business (even if it is small or informal) in this community in the past 2 years?	H3	6.297	0.627	0.500	0.211	236
29	How many houses and small shops (including tables, boxes and kiosks) are selling packaged goods (like cigarettes, biscuits, etc) inside this community today?	H3	3.737	0.626	0.343	0.070	236
30	In the past 2 years, have you participated in any skills training (bookkeeping, soap-making), adult literacy (learn book) or vocation education courses (carpentry, etc.)?	H3	2.831	0.270	0.629	0.668	236
H4: GoBifo increases collective action and contribution to local public goods.							
31	Does this community have any communal farms?	H4	0.144	0.087	0.049	0.074	236
32	Does the primary school that children in the community attend have community	H4	0.746	0.066	0.049	0.180	236
33	Average quality of proposal as assessed by experts	H4	55.309	3.247	1.807	0.074	232
34	Do any people from different households here come together to sell agricultural goods or other petty trading as a group to markets outside of this village (i.e. heap the goods together and send one person to sell; NOT every person totes their own load)?	H4, H7, H8	0.347	-0.046	0.053	0.391	236
35	Average quality of proposal as assessed by policy makers	H4	51.262	2.461	1.591	0.123	232
36	Average completeness of proposal	H4	10.026	-0.013	0.283	0.964	232
37	Whether the proposal is among the top 20 and a winner (as ranked by the Gobifo staff ar	H4	0.093	-0.008	0.037	0.819	236
38	Do any disabled people hold leadership positions in this community (like member of VDC, youth leaders, headman, women's leader, secret society head)?	H5	0.144	0.033	0.048	0.500	236

Appendix Table A5: Raw Results for CDD Effects on Individual Outcomes (continued)

Row	Variable	Hypothesis	Mean, controls	Treatment effect	Standard error	Naïve *p*-value	N
H6: GoBifo changes local systems of authority, including the roles and public perception of traditional leaders (chiefs) versus elected local government.							
55	How old is the current (or acting) village chief/ Headman?	H6	-59.301	-0.974	1.830	0.595	228
56	Enumerator reports on whether "chief decided" project leader nominations	H6	0.873	-0.050	0.043	0.242	235
57	Relative view of "do people in this community believe" Local Councilors as opposed to Chiefdom officials	H6	-0.119	-0.021	0.052	0.684	236
H7: GoBifo increases trust							
58	Are you a member of any credit or savings (osusu) groups?	H7, H8	2.432	0.476	0.285	0.096	236
59	In general, do people in this community believe the central government officials or do they think you need to be careful when dealing with them?	H7	0.314	0.013	0.051	0.794	236
60	In general, do people in this community believe chiefdom officials or do you have to be careful when dealing with them?	H7	0.195	0.053	0.048	0.274	236
61	In general, do people in this community believe Local Councillors or do you have to be careful when dealing with them?	H7	0.076	0.032	0.037	0.392	236
62	In general, do people in this community believe NGOs / donor projects or do you have to be careful when dealing with them?	H7	0.500	0.168	0.057	0.004	236
63	In general, do people in this community believe people from outside you own village / town / neighborhood or do you have to be careful when dealing with them?	H7	0.127	0.088	0.047	0.063	236
64	In general, do people in this community believe people from you own village / town / neighborhood or do you have to be careful when dealing with them?	H7	0.703	-0.069	0.057	0.225	236
H8: Gobifo builds and strengthens community groups and networks							
65	Are there any fishing groups / cooperatives in this community?	H8	0.246	0.037	0.042	0.381	236
66	How many active school PTA groups are there in this village?	H8	4.076	0.719	1.208	0.552	236
67	How many active religious groups (not just going to church/mosque) are there in this village?	H8	4.102	1.721	2.019	0.395	236
68	How many active groups for saving for special events (weddings, funerals) are there in this village?	H8	0.517	0.164	0.116	0.157	236
69	How many active seed multiplication groups are there in this village	H8	0.254	0.853	0.485	0.080	236
70	How many active social clubs are there in this village?	H8	1.441	0.183	0.164	0.265	236
71	How many active women's groups (general) are there in this village?	H8	0.983	-0.039	0.124	0.750	236
72	How many active youth groups (general) are there in this village?	H8	1.212	0.013	0.110	0.907	236

Appendix Table A5: Raw Results for CDD Effects on Individual Outcomes (continued)

Row	Variable	Hypothesis	Mean, controls	Treatment effect	Standard error	Naïve *p*-value	N
H5: GoBifo increases inclusion and participation in community planning and implementation, especially for poor and vulnerable groups; GoBifo norms spill over							
39	Did any disabled people (blind, polio, amputee, wheelchair, etc.) attend the last community meeting?	H5	0.398	0.102	0.063	0.105	236
40	In the past one year, have you attended any community meetings?	H5	-28.644	7.510	7.084	0.290	236
41	Enumerator record of total women (18+ years) present at community meeting	H5	13.264	-0.570	1.176	0.628	216
42	Enumerator record of total youths (18-35 years) present at community meeting	H5	6.009	-0.402	0.391	0.304	229
43	Did anyone take minutes (written record of what was said) at the most recent community meeting?	H5	0.220	0.075	0.056	0.182	236
44	Less concentrated deliberation in manager selection	H5	2.892	0.023	0.090	0.798	231
46	Enumerator account of how democratically the group evenutally came to a decision about who the potential project managers ranging from 5 = open discussion followed by group vote to 1 = chief and/or elders decide without other input	H5	3.364	-0.002	0.094	0.982	235
47	Time of deliberation of manager selection process	H5	32.486	53.665	27.838	0.055	210
48	Enumerator record of total public speakers durings selection of potential project managers	H5	43.429	-2.772	2.584	0.285	213
49	Did a vote occur during the project leader nomination discussion	H5, H6	1.929	0.023	0.032	0.464	171
50	Record of total women (18+ years) in "important people" focus group list	H5	8.186	0.192	0.229	0.402	236
51	Enumerator account of how actively women participated in the deliberation on the selection of potential project managers compared to men, ranging from 5 = no difference between women and men to 1 = women not active at all compared to men	H5	2.799	-0.122	0.132	0.357	232
52	Record of total youth (18-35 years) in "important people" focus group list	H5	2.288	-0.209	0.193	0.281	236
53	Enumerator account of how actively youth participated in the deliberation on the selecti	H5	3.035	0.173	0.153	0.260	229
54	Has this community had any problems with financial mismanagement/corruption in the past 2 years?	H5	0.839	-0.020	0.044	0.657	236

Appendix Table A5: Raw Results for CDD Effects on Individual Outcomes (continued)

Row	Variable	Hypothesis	Mean, controls	Treatment effect	Standard error	Naïve *p*-value	N
H9: GoBifo increases access to information about local governance							
73	Supervisor assessment of whether there are any of the following items--awareness campaigns, financial information, development plan, minutes from any meetings, government policies, election information--visible anywhere around the village (i.e. on a notice board, school, clinic, shop, etc.)?	H9	0.117	0.005	0.018	0.805	236
74	Has this community been visited by the Paramount Chief in the past year?	H9	0.127	-0.023	0.040	0.562	236
H10: GoBifo increases public participation in local governance							
75	Did anyone in this community contest the party symbol in the recent local council elections?	H10	0.169	-0.006	0.044	0.899	236
76	Did anyone in this community stand for the most recent paramount chief elections?	H10	0.068	0.032	0.035	0.358	236
77	Did anyone in this community stand for the most recent section chief elections?	H10	0.280	0.016	0.057	0.777	236
78	Did anyone in this community stand for the most recent Ward Development Committee elections or get nominated for WDC?	H10	0.212	-0.011	0.048	0.813	236
H11: By increasing trust, GoBifo reduces crime and conflict in community.							
79	No conflict that respondent needed help from someone outside the household to resolve in the past one year	H11	-10.424	0.520	1.103	0.638	236
80	In the past 12 months, respondent has not been involved in any physical fighting	H11	-0.568	-0.124	0.270	0.647	236
81	In the past 12 months, no livestock, household items or money stolen from the	H11	-12.127	-1.406	1.267	0.268	236
82	During the last 12 months, respondent has not been a victim of witchcraft (juju)	H11	-1.441	0.441	0.351	0.210	236
H12: GoBifo changes political and social attitudes, making individuals more liberal towards women, more accepting of other ethnic groups and "strangers", and less tolerant of corruption and violence.							
83	Is the current (or acting) village chief/Headman a woman?	H12	0.034	-0.010	0.022	0.653	236
84	Is the current (or acting) village chief/Headman less than 35 years old?	H12	0.009	0.034	0.021	0.109	228

Notes: i) specifications include fixed effects for the disctrict council wards (the unit of stratification) and the two balancing variables from the randomization (total households and distance to road) with robust standard errors; iii) "per comparison" p values are appropriate for a priori interest in an individual outcome

Appendix Table A6: CDD Treatment Effects on Ebola Knowledge Items and Response Actions by District

Outcome	Mean, controls	Treatment effect	Standard error	*p*-value	FDR *q*-value
Panel A: Bombali					
Mean Effects Index (all 13 indicators)	0.000	-0.001	0.053	0.992	.
Knowledge Items					
Mean Effects Index (all 9 knowledge items)	0.000	-0.014	0.070	0.838	
Correctly answers "No" to "Can Ebola spread through air?"	0.923	-0.053	0.042	0.205	0.999
Correctly answers "21" to "How many days can it take for the first to symptoms arise?"	0.564	-0.015	0.071	0.829	0.999
Total (of 11 possible) correct answers to questions about how one can get Ebola	5.154	0.101	0.230	0.663	0.999
Knows correct Ebola hotline number	1.000	0.000	.	.	.
Total (of 10 possible) correct answers regarding how to protect yourself against Ebola	5.064	0.101	0.239	0.673	0.999
Correctly answers "No" to "Drinking salt water can help cure Ebola?"	0.962	0.022	0.023	0.350	0.999
Correctly answers "No" to "Drinking chlorine can help cure Ebola?"	1.000	0.000	.	.	.
Correctly answers "No" to "Can someone spread Ebola before they show symptoms?"	0.795	-0.009	0.061	0.880	0.999
Total correct answers (of 14 possible) regarding symptoms of Ebola	7.641	-0.225	0.309	0.466	0.999
Response Actions					
Mean Effects Index (all 4 response actions)	0.000	0.024	0.074	0.747	
Community had an Ebola task force during the Ebola crisis	0.808	0.051	0.057	0.377	0.999
Community created by-laws in relation to Ebola	0.987	0.013	0.013	0.321	0.999
Communities are more likely to go to formal health facilities (nurse, clinic)	0.949	0.010	0.025	0.686	0.999
Communities are more likely to go to formal health facilities for Ebola (nurse, clinic)	0.962	-0.037	0.037	0.315	0.999
Observations	156				
Panel B: Bonthe					
Mean Effects Index (all 13 indicators)	0.000	0.109**	0.053	0.043	.
Knowledge Items					
Mean Effects Index (all 9 knowledge items)	0.000	0.069	0.073	0.352	0.214
Correctly answers "No" to "Can Ebola spread through air?"	0.725	0.088	0.083	0.291	0.990
Correctly answers "21" to "How many days can it take for the first to symptoms arise?"	0.875	0.070	0.062	0.263	0.990
Total (of 11 possible) correct answers to questions about how one can get Ebola	5.350	-0.179	0.328	0.587	0.990
Knows correct Ebola hotline number	1.000	0.000	.	.	.
Total (of 10 possible) correct answers regarding how to protect yourself against Ebola	4.800	-0.333	0.371	0.373	0.990
Correctly answers "No" to "Drinking salt water can help cure Ebola?"	0.950	0.047	0.033	0.160	0.990
Correctly answers "No" to "Drinking chlorine can help cure Ebola?"	1.000	-0.027	0.027	0.326	0.990
Correctly answers "No" to "Can someone spread Ebola before they show symptoms?"	0.500	0.100	0.098	0.312	0.990
Total correct answers (of 14 possible) regarding symptoms of Ebola	6.525	-0.169	0.322	0.601	0.990
Response Actions					
Mean Effects Index (all 4 response actions)	0.000	0.181*	0.093	0.056	0.127
Community had an Ebola task force during the Ebola crisis	0.375	0.121	0.109	0.270	0.593
Community created by-laws in relation to Ebola	0.750	0.099*	0.051	0.057	0.296
Communities are more likely to go to formal health facilities (nurse, clinic)	0.875	0.030	0.076	0.698	0.985
Communities are more likely to go to formal health facilities for Ebola (nurse, clinic)	0.825	0.063	0.070	0.372	0.593
Observations	80				

*Note: i) significance levels based on naive p-values and indicated by * p<0.10, ** p<0.05, ***p<0.01. ii) specifications include strata for geographic ward and two balancing variables (distance to road and community size) from the randomization; iii) robust standard errors; and iv) this table includes 13 of 15 pre-specified primary outcomes in our PAP, excluding 2 outcomes that are observed for fewer than 20 communities in the data.*

Appendix Table A7: Expert Priors Grant Competition Entry

	Mean	SD
	(1)	(2)
CDD Control, Chiefly Default	35.54	23.03
CDD Control, Technocratic Selection	44.06	22.33
CDD Control, Technocratic Selection & Training	53.63	23.52
CDD Treatment, Chiefly Default	42.20	21.10
CDD Treatment, Technocratic Selection	53.90	20.70
CDD Treatment, Technocratic Selection & Training	65.52	20.89

Notes: Expert priors for each treatment arm, which is CDD crossed with the Technocratic Selection and Training treatment arms, where the latter two arms are part of a companion experiment that was overlaid on the CDD study sampling frame and is analyzed in Casey et al. (forthcoming).

Appendix C: Pre-analysis Plan

We include below the relevant sections of our pre-analysis plan. The plan in its entirety, with time stamps, can be found in the American Economic Association's registry for randomized control trials (https://www.socialscienceregistry.org/trials/1784), where detailed Excel sheets listing all outcome variables (referenced as "PAP Sheets 1, 2, 3 and 4") are also available for download.

Pre-analysis Plan: Two Approaches to Community Development

10 March 2017

PIs: K. Casey, R. Glennerster, E. Miguel and M. Voors

Overview

This research project has four main components. The first evaluates the long run effects of a community driven development (CDD) program in Sierra Leone. The project devolved financial and implementation control over public services to communities, accompanied by intensive social facilitation. The second assesses a low cost technocratic alternative that identifies and supports high competence community members to take better advantage of development opportunities. It leverages local talent, addresses information barriers, and augments existing managerial capital with basic training in project management. A third component elicits expert beliefs about the efficacy of these two approaches and assesses their forecast levels and accuracy. A fourth line of inquiry examines whether participation in CDD affected community response to the Ebola crisis.

Registration timeline

We registered this study with the American Economic Association (AEA) Randomized Control Trial Registry on 16 November 2016. Our trial entry can be found here: http://www.socialscienceregistry.org/trials/1784. On 17 November 2016, we uploaded a data management plan that outlines who would have access to data when, and commits all PIs to not access any data with identifying information until after this PAP is lodged. Fieldwork commenced on 18 November 2016. Our Field Manager Angelica Eguiguren at IPA Sierra Leone was the only person who had access to the data at all times. She uploaded the data to a secure server and will invite the PIs to that dropbox as soon as the PAP is lodged. We lodged an email confirming PI adherence to the data management plan on 9 March 2017. We lodged this PAP on 10 March 2017. We have received IRB clearance from Stanford (#38846), the Government of Sierra Leone, Office of the Sierra Leone Ethics and Scientific Review Committee (3-11-2016, Wageningen (18-11-2016), Berkeley (2016099099) and MIT (#1612798296) for this trial.

Part I: Long run effects of CDD

Component Overview: Community Driven Development (CDD) is a participatory approach popular with foreign aid donors that involves communities directly in the financial management and implementation of local public goods. CDD has two main aims: i) improve the stock and quality of local public goods via the provision of block grants; and ii) democratize local decision-making via intensive social facilitation focused on the participation of marginalized groups.

In earlier work, we analyzed the medium run effects of the "GoBifo" CDD project in Sierra Leone (Casey, Glennerster and Miguel 2012).[1] GoBifo was implemented from 2005 to 2009 and provided roughly $5,000 in block grants and six months of dedicated social facilitation per community. The medium run study found substantial positive impacts on local public goods and economic activity, stronger links between the community and local government, and no evidence for more inclusive local decision-making.

[1] Casey K, Glennerster R, Miguel E (2012) Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan. Quarterly Journal of Economics 127 (4): 1755-1812.

During late 2016, we revisited the 236 communities in the original study to assess long term impacts. In the interim, 60 of the treatment communities received additional support from the GoBifo project. Specifically, these 60 communities received $1,300 for youth empowerment programs in 2010. We do not know how exactly the project management staff selected these 60 communities from the pool of 118 treatment communities, but it was not via random assignment.

Hypotheses: The 12 research hypotheses grouped into two families remain the same as those used in the earlier study.

- Family A of hardware outcomes: "GoBifo creates functional development committees" (H1); "Participation in GoBifo improves the quality of local public services infrastructure" (H2); and "Participation in GoBifo improves general economic welfare" (H3).

- Family B of software outcomes: "Participation in GoBifo increases collective action and contributions to local public goods" (H4); "GoBifo increases inclusion and participation in community planning and implementation, especially for poor and vulnerable groups; GoBifo norms spill over into other types of community decisions, making them more inclusive, transparent and accountable" (H5); "GoBifo changes local systems of authority, including the roles and public perception of traditional leaders (chiefs) versus elected local government" (H6);[2] "Participation in GoBifo increases trust" (H7); "Participation in GoBifo builds and strengthens community groups and networks" (H8); "Participation in GoBifo increases access to information about local governance" (H9); "GoBifo increases public participation in local governance" (H10); "By increasing trust, GoBifo reduces crime and conflict in the community" (H11); and "GoBifo changes political and social attitudes, making individuals more liberal towards women, more accepting of other ethnic groups and 'strangers', and less tolerant of corruption and violence" (H12).

Econometric Specifications: For Part I, the primary test of interest is evaluating long run effects of CDD at the family level. Our core specification evaluates treatment effects for Family A and B, using the following model:

$$Y_c^L = \beta_0 + \beta_1 T_c + X'_c \Gamma + W'_c \Pi + \varepsilon_c \quad (1)$$

where Y_c^L is the mean index for each family for community c in the 2016 survey round; T_c is the GoBifo treatment indicator; X_c contains two village-level balancing variables from the randomization process (distance from a road and total number of households); W_c is a fixed effect for geographic ward, the administrative level on which the randomization was stratified; and ε_c is the usual idiosyncratic error term. The parameter of interest is β_1, the average long run treatment effect. We will construct mean effects indices following Kling, Liebman and Katz (2007).[3]

To interpret these effects, we will test whether long run effects differ from the medium run effects in areas where the medium run effects were nonzero (Family A). Here we will test for decay using the following model:

$$Y_c^L - Y_c^M = \gamma_0 + \gamma_1 T_c + X'_c \Lambda + W'_c \Theta + \mu_c \quad (2)$$

[2] As before, that this is not an explicit objective of the GoBifo project leadership itself, but is a plausible research hypothesis.
[3] Kling, J., J. Lieberman and L. Katz (2007) Experimental Analysis of Neighborhood Effects, Econometrica, 75(1); 83–119

where the dependent variable is the difference in mean effects indices measured in the 2016 survey, Y_c^L, and 2009, Y_c^M. The coefficient of interest is γ_1, where $\gamma_1 < 0$ suggests that the treatment effect has dissipated over time for that hypothesis. A combination of failing to reject $\beta_1 = 0$ while rejecting $\gamma_1 \geq 0$ suggests that previously observed treatment effects have dissipated, while failing to reject $\beta_1 = 0$ and $\gamma_1 \geq 0$ presents a less conclusive middle ground that likely reflects greater noise in measuring long run outcomes and accompanying reductions in the power to detect treatment effects. Note that the exact set of outcomes varies between the 2009 and 2016 data collection rounds, so each index will incorporate the relevant outcomes for that particular survey round (see below).

The second test of interest is running Equations (1) and (2) at the hypothesis level where Equation (2) will again only be run for hypotheses with non-zero medium run effects.

Throughout our analysis, we will adjust for the fact that we are running more than one test on the same dataset by implementing false discovery rate (FDR) corrections. Research practice appears to be moving towards FDR and away from the more conservative familywise error rate (FWER) corrections where there are several tests of interest. Since our earlier paper used FWER corrections, we will also report them here to maintain consistency, but note that the preferred specifications use FDR. These adjustments run across the two families (Family A and Family B) or 12 hypotheses (H1 – H12) as relevant. See Benjamini, Krieger and Yekutieli (2006) and Anderson (2008).[4] For all tests, we will also report the "naïve" or "per comparison" *p*-value.

Our third test of interest highlights a few individual outcome measures from a new structured community activity (SCA). Here we will test for long run effects of GoBifo on the managerial capital of community members and the quality of proposals submitted to a project challenge competition run by the local District Councils (discussed in greater detail below). These outcomes measure whether the learning-by-doing experience of participating in GoBifo translates into long run differences in ability to act collectively and take advantage of development opportunities. We will test them as part of our larger research framework under H1 and H4, respectively, but also highlight them on their own as they capture an important channel through which GoBifo could lead to long run changes.

To further interpret the family- and hypothesis-level results, we will also estimate Equation (1) at the level of individual outcome (adjusting for FDR across all outcomes under a given hypothesis). Note that this reporting of all individual outcomes is for illustrative and interpretation purposes only.

Measurement and survey instruments: See ["SES - Endline 2016"]. The main data collection instrument for the long run effects closely follows the community modules used in the 2009 survey. This includes a focus group discussion with local leaders and enumerator physical inspection of community amenities and market activity. Where possible, we have included a community-level analogue of household level indicators included in the 2009 survey. In addition to economic and social outcomes, we include measures of institutional outcomes using the new project challenge SCA. These are captured in several instruments ["Managerial capital test", "Manager selection tally sheet enumerator A and B", "Submission survey", "Submission form", "Technical scoring", "Policy Scoring", "Expert Scoring"]. We did not repeat the household level survey due to budget constraints.

[4] Benjamini, Y., A. Krieger, and D. Yekutieli (2006) Adaptive Linear Step-Up Procedures That Control the False Discovery Rate, Biometrika, 93: 491–507. Anderson, M (2008) 'Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedaian, Perry Preschool, and Early Training Projects,' Journal of the American Statistical Association, 103 (484): 1481–1495.

Outcomes: See ["PAP, sheet 1"]. The table maps each individual outcome to the hypothesis of interest. To facilitate comparison to our earlier work, the first several columns of this table reproduce exactly those in the Appendix J: Raw Results from the supplementary materials to the 2012 QJE article. The list of outcomes has evolved in a few key ways. First, the present data collection uses only community modules and does not conduct household visits. Thus, all household level outcomes (indicated by "HH" in column K "2009 survey level") are omitted. Where possible, we have included a community-level analogue in the current survey (see column O "Additional question 2016"). Second, we exclude almost all conditional outcomes (i.e. those that are contingent on having a specific good in the community) that are only observed for a subset of villages. Third, as part of our new SCA, we designed measures that mirror some of the process-oriented 2009 SCA outcomes (e.g. unobtrusively counting the number of women who participate in a community decision).

The Casey et al (2012) paper included 334 outcomes, excluding the conditional variables a total of 206 variables remain (see Table 2 in the paper). The 2016 survey round includes 101 outcomes. Table 1 displays the number of outcomes by hypothesis. In total, 96 outcomes exactly match across both rounds. As a robustness analysis, we rerun Equation (1) and Equation (2) for both survey rounds at the family level restricting the analysis to the 96 variables that appear in both 2009 and 2016 survey rounds.

Table 1. Non-conditional outcomes by Hypothesis

Hypothesis	2009	2016	Matching outcome in both rounds
Family A			
H1	7	6	5
H2	18	17	17
H3	15	7	7
Family B			
H4	15	10	6
H5	47	19	19
H6	25	4	4
H7	12	8	8
H8	15	9	9
H9	17	4	4
H10	18	9	9
H11	8	4	4
H12	9	4	4
Total	**206**	**101**	**96**

Heterogeneous Treatment Effects: We will test for heterogeneous treatment effects along the same eight community-level dimensions we used (and measured) in our earlier analysis (total households, war exposure, average schooling, distance to road, historical domestic slavery, district, ethnic fractionalization and chiefly authority). As an exploratory exercise, we will use an automated process (LASSO and BART) to identify other dimensions that are correlated with heterogeneous effects to mine the data in a principled way.

Part II: Managerial Capital

[PART II IS OMITTED HERE AS IT COVERS THE ANALYSIS IN OUR COMPANION PAPER – see Casey, Glennerster, Miguel and Voors (forthcoming)]

Part III: Expert Beliefs

Component Overview: There have now been several randomized control trials of CDD projects in different countries, most of which find some positive impacts on economic outcomes and little effect on institutions. A key unanswered question is whether experts—in academia and more importantly in policy—are updating their beliefs about how effective CDD projects are. This is important in light of the large amounts of foreign aid at stake ($85 billion spent on CDD in about two decades by the World Bank alone, according to Mansuri and Rao 2012), and whether the accumulation of evidence impacts the allocation of donor funds. We surveyed students, academic and policy experts to elicit their beliefs (following DellaVigna and Pope 2016) about the long run effects of the Sierra Leone CDD project and to forecast how well communities will perform in the new project competition.[5]

We fielded this survey among several distinct groups of experts: i) policy makers working for multilateral aid agencies (including the World Bank, DfID, UNDP and IRC); ii) policy makers in Sierra Leone with knowledge of the GoBifo project; iii) economics graduate students in the US (at UC Berkeley) and the Netherlands (at Wageningen University); iv) economics undergraduate students in Sierra Leone (Fourah Bay College), v) researchers directly involved in evaluating CDD projects other development (economics) researchers; and vi) the PIs of this study. There were two versions of the survey: version 1 provided detailed information on our medium run results and version 2 asked the respondent to make predictions without any information provided. For the majority of respondents, we randomized whether they completed version 1 or 2. A small subset completed both versions.

Hypotheses:

- Estimated long run treatment effects are not the same as the average prior beliefs of surveyed experts (H-III.1)
- Average prior beliefs and forecast accuracy differ across groups of experts (H-III.2)
- Prior beliefs about long run effects of the GoBifo project are more optimistic (e.g. predict larger positive long run effects) amongst policy makers compared to researchers (H-III.3)
- Predictions under version 1 of the survey (that contains information on the medium run effects) are more accurate than under version 2 (H-III.4)

Econometric Specifications: For Hypothesis H-III.1, we will evaluate whether the average prior belief across all six groups of experts are statistically distinguishable from the estimated long run treatment effects by GoBifo family and hypothesis. For H-III.2 we will test whether mean predicted effect size by family varies across groups, and assess which estimate is closest to the observed long run effects. H-III.3 tests whether the mean prior of expert groups i and ii more optimistic (predict large positive effects) than that of groups v and vi, at the family level (one sided test). Tests of H-III.4 whether prior beliefs are more accurate in version 2 compared to version 1 across all six groups. For H-III.4 we will use all the data. As a robustness check we will drop data from the subset of respondents that completed both versions of the survey.

We will run several additional descriptive analyses. These include testing whether respondents who report higher confidence in their estimates, and greater familiarity with the 2012 study, are more accurate in their

[5] DellaVigna, S. and D. Pope, "Predicting Experimental Results: Who Knows What?" NBER Working Paper No. 22566, August 2016. See also Humphreys, M., R. Sanchez de la Sierra and P. van der Windt (2016) Social Engineering in the Tropics: A Grassroots Democratization Experiment in Congo, working paper.

predictions. For the new SCA project challenge, we will impute several estimates—regarding GoBifo treatment effects, the efficacy of training, and the impact of technocratic manager selection—and compare their mean values and accuracy across expert respondent groups.[6]

Measurement and Survey Instruments: See ["Expert Priors Survey"]

Outcomes: See ["PAP, sheet 3"].

Part IV: Impacts on Ebola

Component Overview: The recent outbreak of Ebola Virus Disease (EVD) in West Africa is the largest ever recorded. The crisis resulted in over 4000 deaths in Sierra Leone alone (about 11000 in total). The two districts where GoBifo was implemented were differentially effected, Bombali saw 1050 suspected cases and 391 deaths, while Bonthe was much less hit, with 5 suspected cases and 5 deaths. In addition to Communities suffered directly due to fear, illness and loss of life, and indirectly due to travel and trade restrictions resulting from imposed quarantines. The Ebola crisis provided a huge stress on communities at social, political and economic levels. We analyze if participation in Gobifo put communities in a better position to implement preventative measures and collaborate with local government. We report two secondary outcomes (i) we separate impacts on knowledge and collective action, and (ii) we investigate if Gobifo villages reported different Ebola case-loads.

Hypothesis: Our main hypothesis is that "Participation in GoBifo increased knowledge, collective action and investments in preventative measures during the Ebola crisis".

Econometric Specifications: same as Equation (1) above. Our dependent variable is a mean effects index of all Ebola related outcomes. As secondary outcomes, we assess impacts in a mean effects index for knowledge and collective action outcomes separately.

We asses outcomes for the whole sample and restrict our sample to Bombali, which saw many more Ebola cases than Bonthe making the collective action outcomes more relevant.

To further interpret the hypothesis-level results, we will also estimate Equation (1) at the level of individual outcome, adjusting for FDR across outcomes. Note that this reporting of all individual outcomes is for illustrative and interpretation purposes only.

Measurement and survey instruments: see ["SES - Endline 2016", module J and K].

Outcomes: See ["PAP, sheet 4"].

[6] We exclude the study PIs (group vi) from this comparison. While the PIs had no access to the data, we did learn through communication with the field team that the number of submitted proposals was very high.