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

Component Overview: To evaluate a technocratic alternative to CDD's intensive social facilitation model, we overlaid a new randomized experiment across the GoBifo treatment arms. We will test whether i) a more technocratic approach to identifying project leaders with high managerial capital, and ii) the provision of training in project management fundamentals, improves community ability to active collectively and take advantage of a new development opportunity. Specifically, all communities had an opportunity to

enter a project challenge competition run by the local District Councils that awarded US$2,000 implementation grants to the twenty best project proposals. We block randomized 80 communities to a management selection treatment arm (*MS*); 78 to a management selection plus training arm (*MST*); and 80 to a control or status quo (*SQ*) mechanism that favors the village headmen.

These three treatment arms were implemented by the research team enumerators on the data collection visits to communities at the end of the focus group discussion. In all three arms, enumerators explained the project challenge opportunity and the skills needed to develop a strong proposal. They asked the group to deliberate and nominate five individuals, in addition to the village headman, who had these skills. These 6 individuals were then asked to take a management test, in private, which was scored on site by enumerators. The focus group was then reconvened and a public lottery (implemented on a tablet device) determined treatment assignment for the village. In the *status quo (SQ) arm*, the village headman was designated as the project proposal leader. His name was written on the standardized project application form and he was given a transportation voucher to redeem if/when he submitted a proposal to the relevant Local Council. In the *manager selection (MS) arm*, the enumerators announced who was the highest test performer (of the 5 non-chief nominees), and designated that person on the submission form and provided the transport voucher. The *manager selection plus training (MSTR) arm* followed the same format as *MS* but also announced that the relevant ward development committee (most local tier of elected government) would hold a one day management training as part of the project challenge competition. Enumerators provided the date and location of the training, informed the group that the travel costs of the designated project leader will be reimbursed, and encouraged the designated project leader to attend the training.

The training sessions for *MSTR* covered: i) identification of local development needs and designing projects to address them; ii) costing local materials and developing itemized budgets; and iii) time management and planning to meet deadlines. Note that measures of proposal quality capture both items covered in the training and those that were not, to evaluate the extent to which any observed training effects reflect "teaching to the test."

Hypotheses: We plan to evaluate the following hypotheses:

- There is underutilized managerial capital in villages (H-II.1)
- Leveraging underutilized managerial capital leads to greater ability to act collectively and take advantage of local development opportunities (H-II.2)
- Lack of management skills constrains the ability to take advantage of local development opportunities (H-II.3).

Econometric Specifications: Our primary tests of interest estimate:

$$P_c = \delta_0 + \delta_1 MS_c + \delta_2 TR_c + W'_c \Psi + \zeta_c \quad (3)$$

where outcome *P* (i.e. proposal quality, test score of project leader) is measured for community *c*; *MS* is an indicator variable equal to one for assignment to the manager selection process (*MS* and *MSTR* arms) and zero otherwise; TR is an indicator for assignment to training (*MSTR* arm); W_c is a stratification fixed effect for geographic wards; and ζ_c the idiosyncratic error term. Hypotheses H-II.1 and H-II.2 test $\delta_1 = 0$. Hypothesis H-II.3 tests $\delta_2 = 0$.

For Hypothesis H-II.1 we have only one outcome, the test score of the project proposal leader. For Hypotheses H-II.2 and H-II.3 we have four measures of proposal quality so our primary specification will

be a mean effects index. We will also report estimates for the individual scores. As a robustness check, we will exclude quality assessments that involve any input from GoBifo staff (although note all proposals were blinded during the review).

Several additional analyses will aid in interpreting these results (see [PAP Sheet 2] for details). We will:

1. Explore the extent to which the training reflects "teaching to the test." Explore where the training appears most effective.
2. Validate the management test by correlating test scores with proposal quality and explore relative predictive of power of subsection scores.
3. Validate the extent to which the distinct manager selection treatment arms translated into differences in who actually managed the project proposal process.
4. Compare the tests scores of the non-headman nominees to those of village headmen.
5. Evaluate which characteristics correlate with managerial capital test scores (i.e. age, gender, education, management experience, leadership position, etc.).
6. Test for heterogeneous response to training by management test score.
7. Test for interaction effects between participation in GoBifo and the *MS* and *TR* terms in Equation 3, noting that these tests are likely underpowered.

Measurement and Survey Instruments: We used several instruments to implement and evaluate this new SCA, see ["Managerial capital test", "Manager selection tally sheet enumerator A and B", "Submission survey", "Submission form", "Technical scoring", "Policy Scoring", "Expert Scoring" and data from the transcripts of the training].

Outcomes: See ["PAP, Sheet 2"]

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

[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.

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 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.

[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.

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"].