ONLINE APPENDIX: MATERIAL NOT INTENDED FOR PUBLICATION

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

THIS VERSION: 18 October 2022

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.

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 ###-###-####. If you have any complaints, please contact the Stanford Human Subjects Institutional Review Board (IRB) at +1 ###-###-####.

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

    _____

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   *Very familiar*
    with results

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 <u>standard deviation units (sdu's)</u>. 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?


```
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
 <-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*
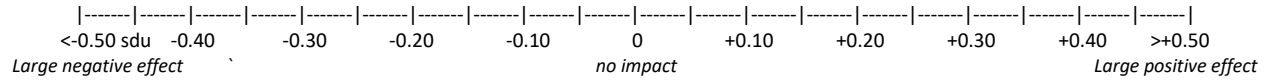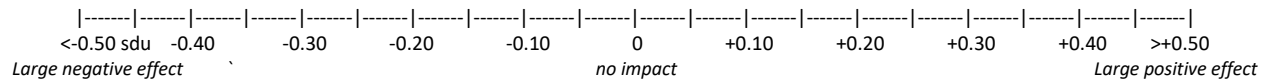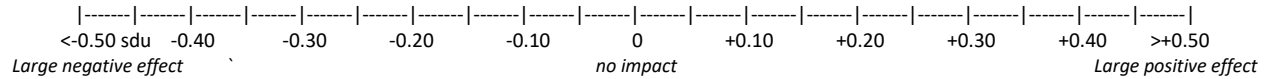
Hypothesis 2: Participation in GoBifo improves the quality of local public services infrastructure.
Examples include the presence and construction quality of latrines and drying floors.

*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?

```
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
 <-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 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?

```
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
 <-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*

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.*

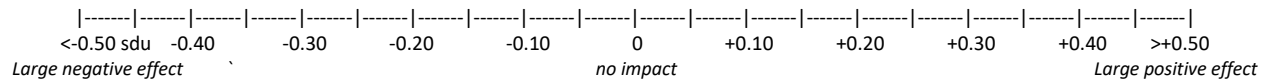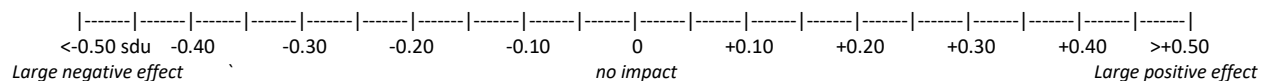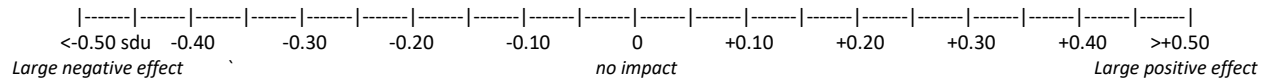What do you think the long run treatment effect will be?

```
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-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.*
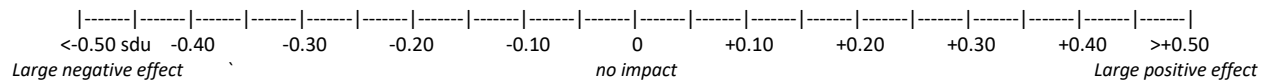
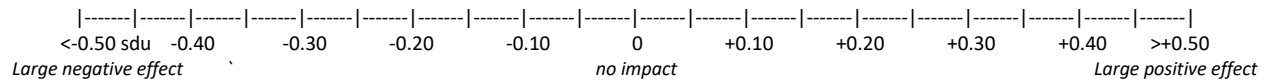What do you think the long run treatment effect will be?

```
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-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.*

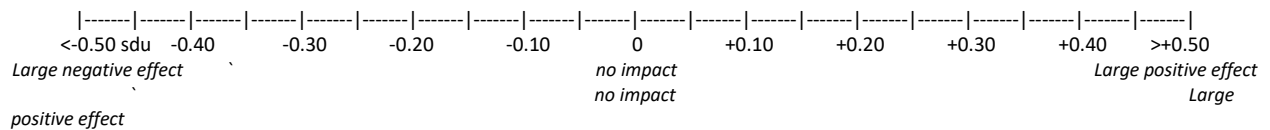What do you think the long run treatment effect will be?

```
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-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.*

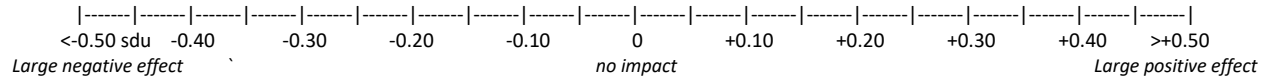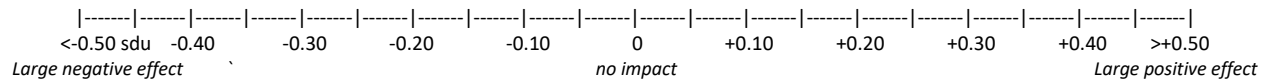What do you think the long run treatment effect will be?

```
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
 <-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 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?

```
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
 <-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 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?

```
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
 <-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 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?

```
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
 <-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*
         `   *no impact*   *Large*

positive effect

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?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-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 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?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-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*

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

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

	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 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)	(0.081)
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 ($N=57$)	0.065**	0.015	0.008	0.086***	-0.020
	(0.026)			(0.030)	(0.033)
Collective action	0.104*	0.053	0.106	0.072	0.032
	(0.053)			(0.046)	(0.065)
Inclusion	0.036	0.338	0.291	0.084*	-0.047
	(0.038)			(0.049)	(0.052)
Local authority	-0.050	0.380	0.296	0.110	-0.160*
	(0.056)			(0.068)	(0.083)
Trust	0.107*	0.065	0.109	0.032	0.074
	(0.057)			(0.049)	(0.083)
Groups and networks	0.149**	0.038	0.094	0.056	0.093
	(0.071)			(0.045)	(0.080)
Access to information	-0.036	0.591	0.476	0.150**	-0.187**
	(0.067)			(0.072)	(0.092)
Participation in governance	0.079	0.191	0.194	0.256**	-0.177**
	(0.060)			(0.058)	(0.068)
Crime and conflict	-0.002	0.971	0.480	0.088	-0.090
	(0.063)			(0.062)	(0.084)
Political and social attitudes	0.154	0.216	0.194	-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) column 3 includes q-values from false discovery rate (FDR) corrections across the 12 hypotheses and across the two family indices respectively (see Benjamini, Krieger and Yekutieli 2006 and Anderson 2008); vi) outcomes limited to those that were collected in the exact same fashion in both 2009 and 2016 survey rounds; and vii) 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 goods

  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 accountable

  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 researchers

  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.

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**	0.270*
	(0.203)	(0.126)
Treatment * Total households in the community	-0.002	-0.001
	(0.002)	(0.001)
Treatment * Index of war Exposure	-0.306	-0.027
	(0.232)	(0.147)
Treatment * Average respondent schooling	0.008	0.005
	(0.035)	(0.021)
Treatment * Distance to motorable road	0.005	0.005
	(0.017)	(0.009)
Treatment * Historical extent of domestic slavery	-0.108	-0.054
	(0.086)	(0.061)
Treatment * Bombali district	-0.457**	-0.152**
	(0.089)	(0.056)
Treatment * Ethnolinguistic fractionalization	0.122	-0.247
	(0.239)	(0.158)
Treatment * Chiefly authority	0.141	-0.103
	(0.287)	(0.215)
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 11 SCA outcomes			0.000	0.009	0.057	0.876	236
123	Potential managers selection deliberation done in public debate	H5	0.584	-0.024	0.056	0.668	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 eventually 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	0.038	0.022	0.026	0.392	160
201	Enumerator reports on whether "chief decided" project leader nominations	H6	0.876	-0.036	0.038	0.351	231

*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; iii) robust standard errors; and iv) this table includes 11 of 15 pre-specified primary outcomes in our PAP, excluding outcomes corresponding to the grants competition (Casey et al., 2021).*

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
H5:GoBifo increases inclusion and participation in community planning and implementation, especially for poor and vulnerable groups; GoBifo norms							
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	Days since last community meeting	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 eventually 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	0.038	0.022	0.026	0.392	160
50	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
51	Enumerator account of how actively youth participated in the deliberation on the selecti	H5	3.035	0.173	0.153	0.260	229
52	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
H6: GoBifo changes local systems of authority, including the roles and public perception of traditional leaders (chiefs) versus elected local government.							
53	How old is the current (or acting) village chief/ Headman?	H6	-59.301	-0.974	1.830	0.595	228
54	Enumerator reports on whether "chief decided" project leader nominations	H6	0.876	-0.036	0.038	0.351	231
55	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							
56	Are you a member of any credit or savings (osusu) groups?	H7, H8	2.432	0.476	0.285	0.096	236
57	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
58	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
59	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
60	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
61	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
62	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							
63	Are there any fishing groups / cooperatives in this community?	H8	0.246	0.037	0.042	0.381	236
64	How many active school PTA groups are there in this village?	H8	4.076	0.719	1.208	0.552	236
65	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
66	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
67	How many active seed multiplication groups are there in this village	H8	0.254	0.853	0.485	0.080	236
68	How many active social clubs are there in this village?	H8	1.441	0.183	0.164	0.265	236
69	How many active women's groups (general) are there in this village?	H8	0.983	-0.039	0.124	0.750	236
70	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
H9: GoBifo increases access to information about local governance							
71	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
72	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							
73	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
74	Did anyone in this community stand for the most recent paramount chief elections?	H10	0.068	0.032	0.035	0.358	236
75	Did anyone in this community stand for the most recent section chief elections?	H10	0.280	0.016	0.057	0.777	236
76	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.							
77	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
78	In the past 12 months, respondent has not been involved in any physical fighting	H11	-0.568	-0.124	0.270	0.647	236
79	In the past 12 months, no livestock, household items or money stolen from the	H11	-12.127	-1.406	1.267	0.268	236
80	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.							
81	Is the current (or acting) village chief/Headman a woman?	H12	0.034	-0.010	0.022	0.653	236
82	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) significance levels (per comparison p-value) indicated by * p <0.10, ** p <0.05, *** p <0.01; ii) specification that includes 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	0.999
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	0.999
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	0.990
Community created by-laws in relation to Ebola	0.750	0.099*	0.051	0.057	0.990
Communities are more likely to go to formal health facilities (nurse, clinic)	0.875	0.030	0.076	0.698	0.990
Communities are more likely to go to formal health facilities for Ebola (nurse, clinic)	0.825	0.063	0.070	0.372	0.990
Observations	50				

*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) 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; and v) q-values are from false discovery rate (FDR) corrections adjusted across the two category-level indices, or across all 13 individual knowledge and response outcomes, respectively.*

Appendix Table A7: Predicted Entry into Grants Competition by Experimental Arms

2016 New experiment assignment:	2005 CDD assignment:	
	CDD Control	CDD Treatment
Status Quo Chiefs	*Arm 1*	*Arm 4*
	35.5%	42.2%
	(23.0)	(21.1)
Technocratic Selection	*Arm 2*	*Arm 5*
	44.0%	53.9%
	(22.3)	(20.7)
Trained Technocrats	*Arm 3*	*Arm 6*
	53.6%	65.5%
	(23.5)	(20.9)
Realized entry, all communities:	98.3%	

Notes: The expert forecasting survey asked for predictions for each of six experimental treatment arms: these arms include the original treatment and control arms of the long-run CDD program analyzed here, plus an overlay of treatment arms for a new experiment that is analyzed in Casey et al (2021). As a complement to the expert predictions for the pure control arm (Arm 1 above) displayed in main text Figure 6 panel C (which is the cell at the intersection of the CDD control arm and the control arm of the new experiment), this table presents mean expert predictions about the percent of communities that would enter the project challenge competition for all six distinct treatment arms.

Appendix Table A8: Baseline (2005) Balance by Treatment Assignment

	Baseline mean for controls	T-C difference at baseline	N
	(1)	(2)	(3)
Panel A: Community Characteristics			
Total households per community	46.76	0.30	236
		(3.67)	
Distance to nearest motorable road in miles	2.99	-0.32	236
		(0.36)	
Index of war exposure (range 0 to 1)	0.68	-0.01	236
		(0.02)	
Historical extent of domestic slavery (range 0 to 1)	0.36	0.03	236
		(0.06)	
Average respondent years of education	1.65	0.11	235
		(0.13)	
Panel B: Selected Variables from "Hardware" Family A			
Proportion of communities with a Village development committee (VDC)	0.55	0.06	232
		(0.06)	
Proportion visited by Ward Development Committee (WDC) member in past year	0.15	-0.01	228
		(0.05)	
Proportion of communities with a functional grain drying floor	0.23	0.05	231
		(0.05)	
Proportion of communities with a functional primary school	0.41	0.08	230
		(0.06)	
Average household asset score	-0.06	0.11	235
		(0.08)	
Proportion of communities with any petty traders	0.54	-0.01	226
		(0.06)	
Panel C: Selected Variables from "Software" Family B			
Respondent agrees that chiefdom officials can be trusted	0.66	-0.01	235
		(0.02)	
Respondent agrees that Local Councillors can be trusted	0.61	0.00	235
		(0.02)	
Respondent is a member of credit / savings group	0.25	-0.03	235
		(0.02)	
Among males who attended a community meeting, respondent spoke publicly	0.59	-0.02	235
		(0.04)	
Among females who attended a community meeting, respondent spoke publicly	0.29	0.03	229
		(0.04)	
Respondent claimed to have voted in last local elections	0.85	-0.01	235
		(0.02)	

*Notes: i) significance levels indicated by + p <0.10, * p <0.05, ** p <0.01; ii) robust standard errors; iii) the T-C difference is the pre-program "treatment effect" run on the baseline data aggregated to the village-level mean, using a minimal specification that includes only fixed effects for the district council wards (the unit of stratification) and the two balancing variables from the randomization (total households and distance to road); iv) regressions for the two balancing variables in rows 1 and 2 exclude the outcome from the set of controls; and v) this table is reproduced from Casey et al 2012, page 1770.*

Appendix C: Additional Figures

Appendix Figure A1: Persistence of Individual Public Infrastructure Items
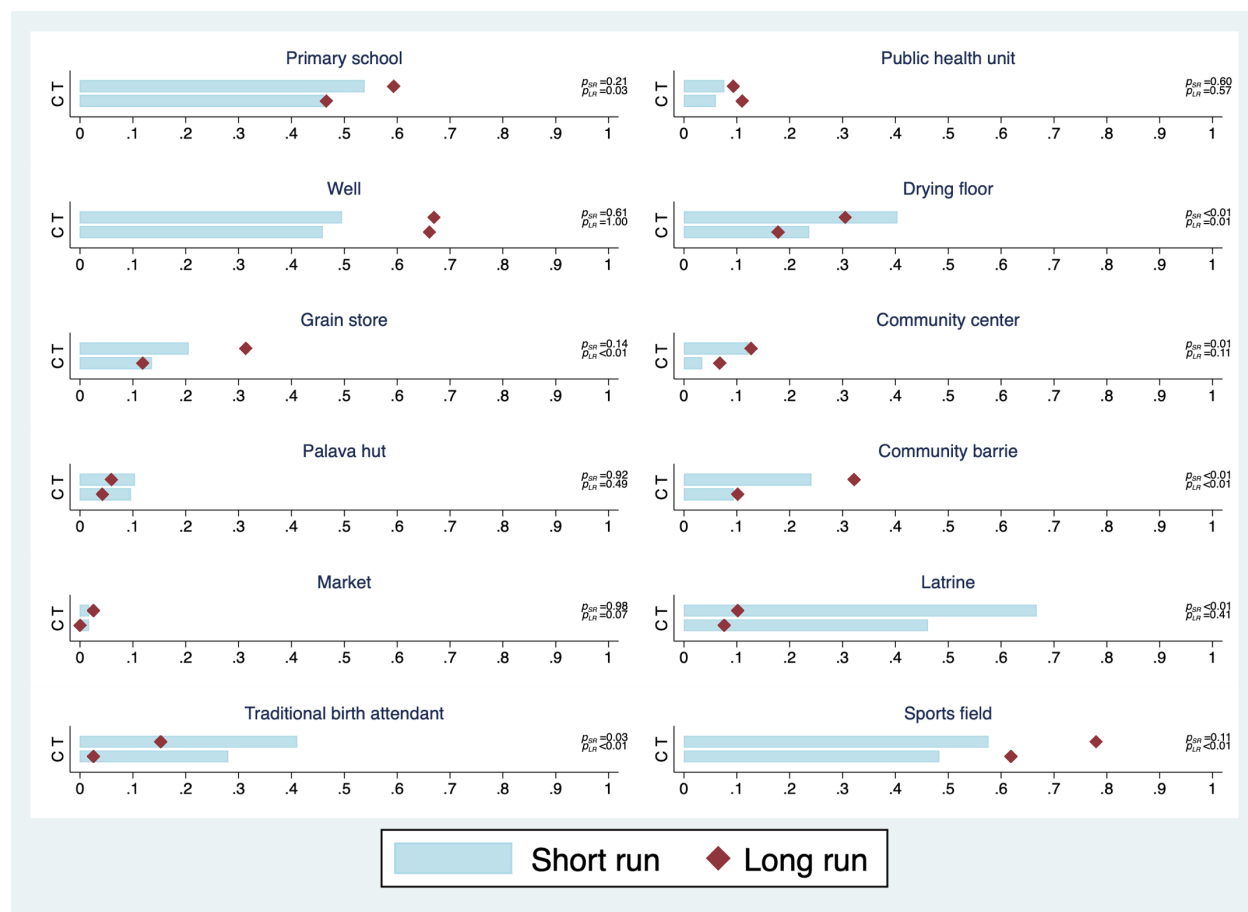
Notes: This figure displays the proportion of communities that have each of 12 local infrastructure items as observed by field enumerator inspection (which are displayed as a group in main text Figure 3). The solid horizontal bars denote the proportion for treatment (top) versus control (bottom) communities observed in the short-run 2009 data. The red diamonds indicate the corresponding proportions observed in the long-run 2016 data. Reported p-values are associated with treatment effect estimates that include the full suite of pre-specified controls in the short-run (indicated "p_{SR}") and long-run ("p_{LR}") datasets, respectively. For traditional birth attendant hut and sports field, the short-run data are aggregated from household reports, while the long-run data are based on enumerator inspection.

Appendix Figure A2: Gains in Individual Economic Welfare Measures over Time
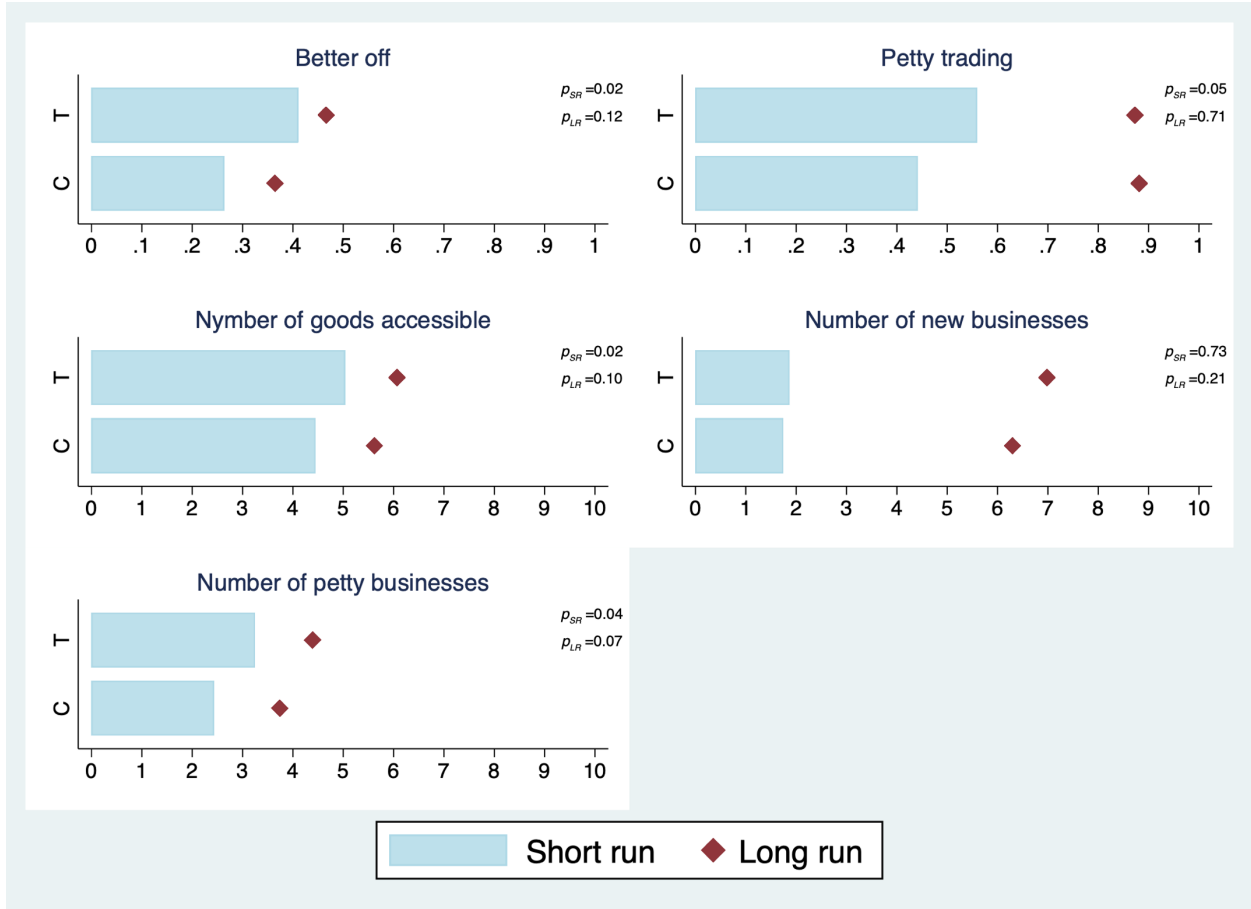
Notes: This figure displays the proportion of communities that have each of 5 economic welfare items (which are displayed as a group in main text Figure 4). The solid horizontal bars denote the proportion for treatment (top) versus control (bottom) communities observed in the short-run 2009 data. The red diamonds indicate the corresponding proportions observed in the long-run 2016 data. Reported p-values are associated with treatment effect estimates that include the full suite of pre-specified controls in the short-run (indicated "p_{SR}") and long-run ("p_{LR}") datasets, respectively

Appendix Figure A3: Correlations across Hardware and Software Families

Panel A: Unconditional Correlation at Endline (2016)
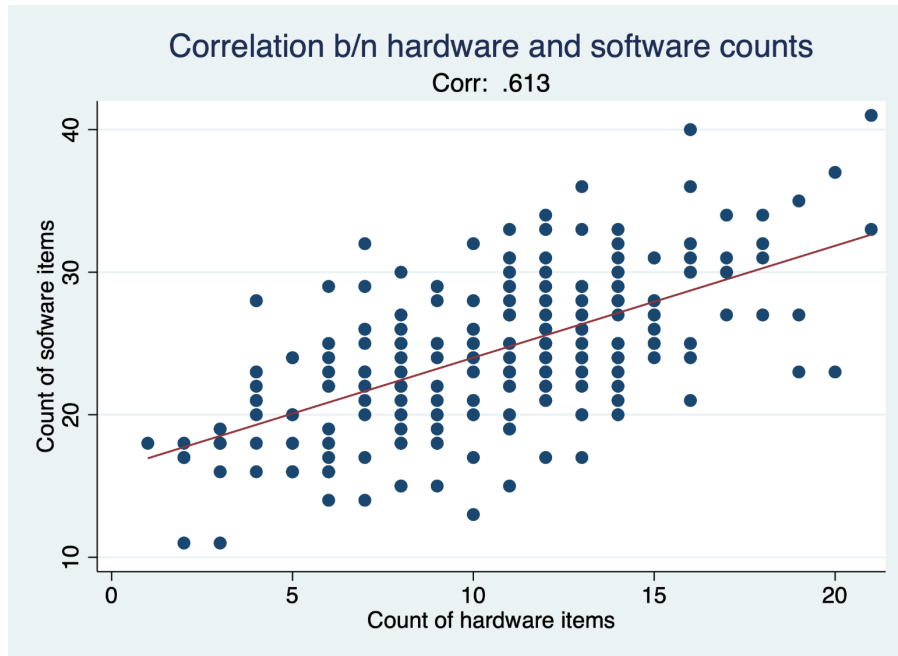
Panel B: Correlation in Changes from 2005 to 2016
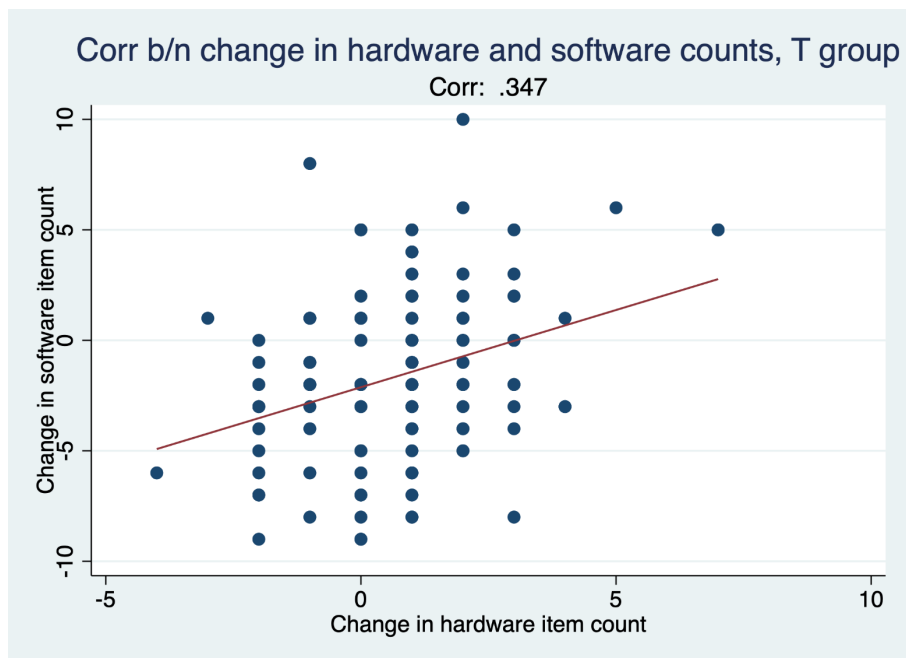
Notes: This figure displays correlations between the two families of outcomes. Panel A correlates the count of 29 infrastructure outcomes with the count of 57 institutional outcomes observed in 2016 (these are the same counts displayed in Figure 5). Panel B correlates the change over time (from baseline in 2005 to 2016) in each of these family-wise counts. There are fewer outcomes that form an exact panel with the 2005 baseline data: 12 (24) for family A (B).

Appendix D: 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.

<div align="center">

Pre-analysis Plan: Two Approaches to Community Development

10 March 2017

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

</div>

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:

$$Y_c^L = \beta_0 + \beta_1 T_c + X'_c \Gamma + W'_c \Pi + \varepsilon_c \tag{1}$$

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:

$$Y_c^L - Y_c^M = \gamma_0 + \gamma_1 T_c + X'_c \Lambda + W'_c \Theta + \mu_c \tag{2}$$

---

[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.