Pre-analysis Plans (PAP's) in Social Science Research: An Application and Discussion

EASST Meeting

Kampala, July 13, 2012

Edward Miguel,

University of California, Berkeley

Experiments and internal validity

- Over the last decade, **field experiments**, lab experiments and other studies featuring original data collection and rigorous research designs (e.g., IV, regression discontinuity, etc.) have become widespread in economics and political science

  -- The talks we heard yesterday are a perfect illustration

- The spread of these tools has been driven by the perception that they have more **internal validity**, and thus credibility, than most observational approaches.

- Even with these gains, where are experimental studies still falling short – and how can we do better? I'll focus on one dimension.

Learning from medical trials

- For the same reasons experiments spread in social science research, **randomized drug trials** started decades ago

- However, they were not without their critics: perhaps because of the massive profits on the line, pharmaceutical companies sometimes **suppressed "failed" trials**, or focused on alternative outcomes ex post in order to promote particular drugs

- The solution: a U.S. National Institutes of Health (NIH) backed medical trial "registry" became standard circa 2000, and most major journals began requiring pre-registration of analysis plans for publication.

Registering pre-analysis plans (PAP's)

- The idea: by requiring PAP registration for funding and journal publication, all trials are placed "in the public domain", allowing for a more complete sense of the results in the literature (for meta-analyses, for instance), and limiting publication bias.

- Pre-specifying both the main outcome variables and statistical approach also helps guard against data mining, specification search (Leamer 1974, 1983), and **"cherry picking"** outcomes or subgroups that have significant impacts (potentially by chance)

- Use of the PAP registry for medical trials has become universal, and is thought to have reduced the worst abuses. Could the use of PAP's have similar benefits for social science research?

An application to political economy

- Casey, Glennerster and Miguel (2012) estimates the impact of a community driven development (CDD) program in post-war Sierra Leone on a range of local public goods outcomes, as well as institutional performance, social capital, and local politics.

- CDD aims to improve the capacity and performance of local village governments (through elected committees), and boost inclusion of marginalized groups, including women and youth.

- Large-scale randomized experiment with N=236 villages, a four year time-frame (2005-2009), detailed data collection.
- Institutions are multi-faceted and there are many possible measures: we have **over 300** such measures!

Motivation for CDD

"Experience demonstrates that by directly relying on poor people to drive development activities, CDD [community driven development] has the potential to make poverty reduction efforts more responsive to demands, more inclusive, more sustainable, and more cost-effective than traditional centrally led programs…achieving immediate and lasting results at the grassroots level." – Dongier et al. (2003), World Bank

Why might Sierra Leone's institutions warrant reform?

- **Legacy of bad governance and corruption in the formal system**
  - President Siaka Stevens abolished local government (1972) and banned rival political parties (1978), abysmal public services
- **The traditional system is (also) dominated by elder male elites**
  - 149 Paramount Chiefs rule for life; come from hereditary ruling houses; and control land, labor and the judiciary outside the capital
  - Women are not even eligible for chieftaincy in most of the country

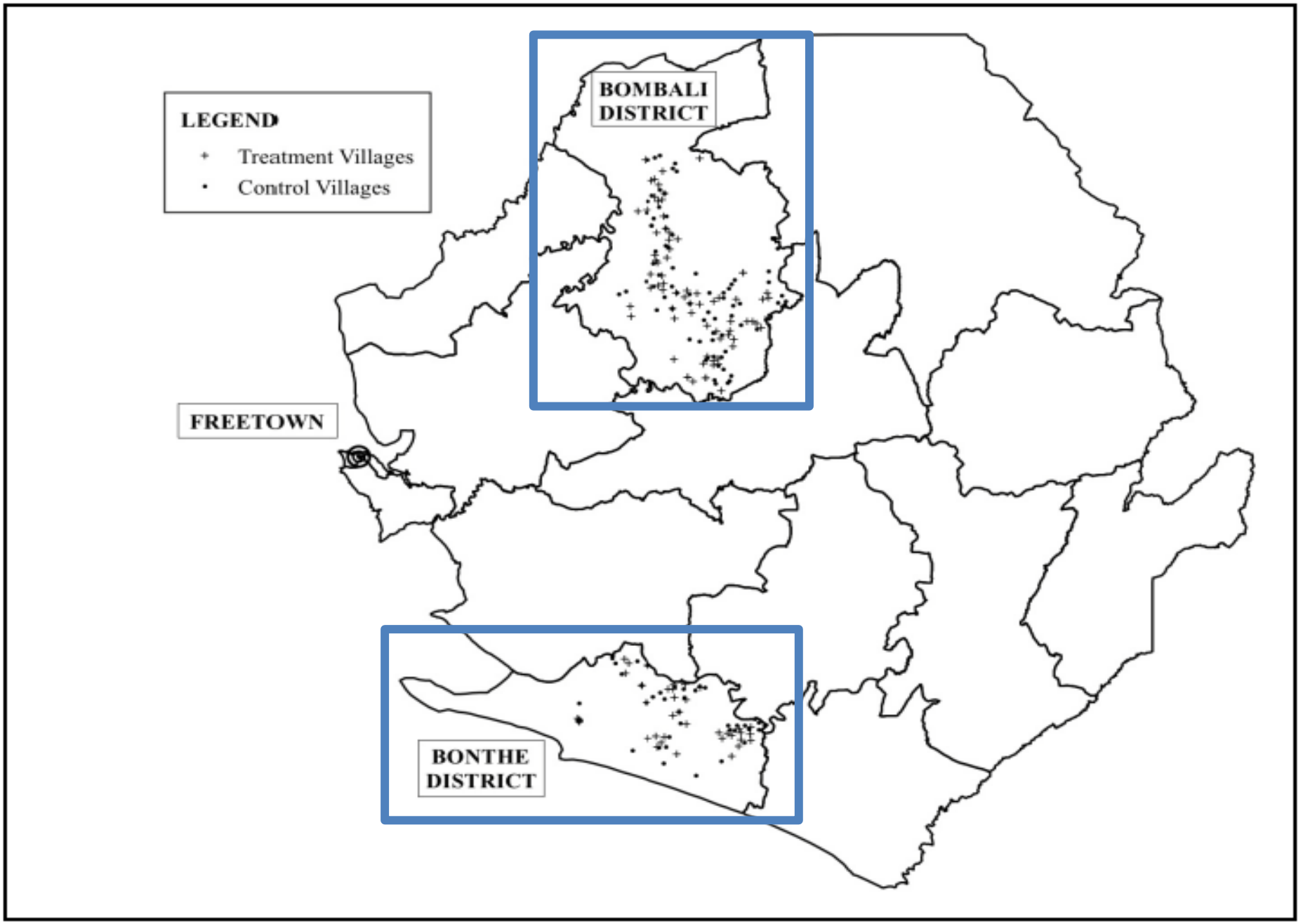
- Scholars point to seeds of the 1991-2002 civil war in social divisions, inequalities, and lack of political representation.

What does CDD aim to do?

- **Financial grants** for local public goods, small enterprise development
  - The "GoBifo" Project ("Move Forward") we study in Sierra Leone gave $4,667 to communities in 3 tranches (~$100 per household)

- **Training and facilitation** to build durable local collective action capacity (6 months of intensive contact spread out over 4 years)
  - Forms a representative Village Development Committee to promote democratic decision-making
  - Establishes bank accounts and transparent accounting procedures
- Requirements to **increase participation of marginalized** groups
  - Women were co-signatories on the community bank accounts
  - Women and youths managed own projects, e.g. labor groups

Appendix D: Location of Research Communities



Local public goods construction projects

- The distribution of community projects by sector was:
  - Infrastructure (43%) - e.g., community centers, primary schools
  - Agriculture/livestock (40%) - e.g., seed multiplication, goats
  - Skills training, small business (17%) - e.g., carpentry, soap-making

Using a PAP

- We wrote up our research hypotheses in October 2005 in collaboration with the project team, and registered a PAP with the Jameel Poverty Action Lab (J-PAL) registry in August 2009 before data analysis (povertyactionlab.org/Hypothesis-Registry).

- Detailed document laying out 11 hypotheses, each with multiple measures; exact econometric specifications, a the mean effects approach used to combine across multiple measures; list of "subgroups" (e.g., by civil war history) to analyze.

Appendix B: Project and Research Timeline

10-Oct-05	*Hypothesis document drafted*
Nov-05	Baseline Survey
Dec-05	
Jan-06	Ward Facilitator Training
Feb-06	
Mar-06	
Apr-06	
May-06	Development Planning
Jun-06	
Jul-06	
Aug-06	
Sep-06	
Oct-06	
Nov-06	
Dec-06	
Jan-07	Ward Development Committee Approval
Feb-07	
Mar-07	
Apr-07	Delays
May-07	
Jun-07	
Jul-07	
Aug-07	
Sep-07	
Nov-07	
Dec-07	
Jan-08	Projects implemented
Feb-08	
Mar-08	
Apr-08	Second grants disbursed
May-08	
Jun-08	Projects implemented
Jul-08	
Aug-08	
Sep-08	Third grants disbursed
Oct-08	
Nov-08	Projects implemented
Dec-08	
Jan-09	
Feb-09	
Mar-09	
Apr-09	
May-09	Follow-up survey 1
Jun-09	Voucher program begins
Jul-09	
21-Aug-09	*Pre-Analysis Plan archived with the Jameel Poverty Action Lab*
Sep-09	Voucher program ends
Oct-09	Follow-up survey 2
Nov-09	
4-Mar-10	*Plan Supplement covering second follow-up survey archived*

Econometric specifications

- Basic model for outcomes with post-program data only:

$$\mathbf{Y_c = \beta_0 + \beta_1 T_c + X_c' \Gamma + W_c' \Pi + \varepsilon_c}$$

  - Y_c is outcome in community c (HH data averaged by village)
  - T_c is an indicator for GoBifo treatment
  - X_c is a vector of community-level controls (pre-specified, results are robust to their exclusion); W_c are ward fixed effects
  - ε_c is an idiosyncratic error term
- Results unchanged with panel specification (where data available)

Using a PAP

- We wrote up our research hypotheses in October 2005 in collaboration with the project team, and registered a PAP with the Jameel Poverty Action Lab (J-PAL) registry in August 2009 before data analysis (povertyactionlab.org/Hypothesis-Registry).

- Detailed document laying out 11 hypotheses, each with multiple measures; exact econometric specifications, a the mean effects approach used to combine across multiple measures; list of "subgroups" (e.g., by civil war history) to analyze.

- Defining hypotheses in advance prevents us from selecting outcomes that tell a great "story", and **shields us from pressure** to report only results that support donor/policymaker agendas.

Overview of results

- **Outcome family A: The project was well-implemented, with strong impacts on "hardware" and economic activity**
  - Village-level structures and tools to manage development projects were established (e.g. bank accounts)
  - Finances were disbursed with little leakage (<13% discrepancies)
  - Increases in the stock and quality of local public goods
  - Increases in household assets and village-level market activity

Table 4: Illustrative Selection of Statistically Significant Treatment Effects, Family A

Outcome Variable	Mean in Controls	Treatment Effect	Standard Error	N
	(1)	(2)	(3)	(4)
Panel B: Hypothesis 2 - Local Public Services				
Functional traditional midwife post in the community	0.08	0.17**	(0.04)	235
Functional latrine in the community	0.46	0.21**	(0.06)	234
Functional community center in the community	0.03	0.09**	(0.03)	236
Community took a proposal to an NGO or donor for funding	0.29	-0.15**	(0.05)	229
Supervisor's physical assessment of construction quality (index from 0 to 1):				
Primary School	0.58	0.11+	(0.06)	123
Grain drying floor	0.38	0.16*	(0.08)	101
Latrine	0.27	0.18**	(0.05)	154
Panel C: Hypothesis 3 - Economic Welfare				
Total petty traders in village	2.43	0.70*	(0.34)	225
Total goods on sale of 10	4.45	0.57*	(0.24)	236
Household asset score	-0.16	0.30**	(0.09)	236
Attended trade skills training	0.06	0.12**	(0.02)	235

Overview of results

- **Outcome family A: The project was well-implemented, with strong impacts on "hardware" and economic activity**
  - Village-level structures and tools to manage development projects were established (e.g. bank accounts)
  - Finances were disbursed with little leakage (<13% discrepancies)
  - Increases in the stock and quality of local public goods
  - Increases in household assets and village-level market activity
- **Outcome family B: Zero impact on "software" / "institutions"**
  - No impacts on participation in decision-making
  - No sustained increase in collective action capacity
  - No change in the "voice" of women and young men
  - Apparent "capture" of new organizations by chiefly authorities
  - Example of communal farms: established but low participation

The results with and without the PAP

- To summarize, the CDD project had positive impacts on local public goods and economic outcomes ("family A" outcomes).
- BUT despite extensive training, facilitation and funding over nearly four years, there were no detectable impacts on any of the institutional, political, or social capital outcomes that we (and the project team) had hypothesized ("family B").

- As you can imagine, some project leaders were not thrilled by the family B results…

- **What would results have looked like without a PAP?**

Illustrating the risk of “cherry-picking”

- Given our large number of outcome measures (over 200 for family B), it is possible to selectively present one subset of outcomes for which CDD had a “positive” impact on institutions, and a second subset of outcomes that show the opposite impact.

- Illustrates some of the value of having a pre-analysis plan in place, to limit tendentious reporting.

Table 5: Erroneous Interpretations under "Cherry Picking"

Survey question	Mean for controls	Treatment effect	Standard error	N	Hypo
	(1)	(2)	(3)	(4)	(5)
Panel A: Institutions "Deteriorated"					
Attended meeting to decide what to do with the tarp	0.812	-0.037+	(0.021)	236	H5
Everybody had equal say in deciding how to use the tarp	0.509	-0.106+	(0.058)	232	H5
Correctly able to name what the tarp was used for	0.589	-0.08+	(0.048)	236	H9
Community used the tarp (verified by physical assessment)	0.897	-0.079+	(0.044)	233	H4
Community can show research team the tarp	0.836	-0.116*	(0.051)	232	H5
Respondent would like to be a member of the VDC	0.361	-0.043*	(0.021)	236	H10
Current (or acting) village chief/Headman is younger than 35	0.044	-0.038+	(0.023)	229	H12
Respondent voted in the local government election (2008)	0.851	-0.036*	(0.016)	236	H10
Panel B: Institutions "Improved"					
Community teachers have been trained	0.471	0.122+	(0.066)	173	H4
Respondent is a member of a women's group	0.235	0.060**	(0.021)	236	H8
Someone took minutes at the most recent community meeting	0.295	0.140*	(0.063)	227	H5
Building materials stored in a public place when not in use	0.128	0.246*	(0.098)	84	H5
Chiefdom official did not have the most influence over tarpaulin use	0.543	0.058*	(0.029)	236	H6
Respondent agrees with "Responsible young people can be good leaders" and not "Only older people are mature enough to be leaders"	0.762	0.038*	(0.017)	236	H6, H12
Correctly able to name the Section Chief for this section	0.533	0.053+	(0.032)	234	H9
Correctly able to name the year of the next general elections	0.192	0.038*	(0.018)	236	H9

Table 5: Erroneous Interpretations under "Cherry Picking"

Survey question	Mean for controls	Treatment effect	Standard error	N	Hypo
	(1)	(2)	(3)	(4)	(5)
Panel A: Institutions "Deteriorated"					
Attended meeting to decide what to do with the tarp	0.812	-0.037+	(0.021)	236	H5
Everybody had equal say in deciding how to use the tarp	0.509	-0.106+	(0.058)	232	H5
Correctly able to name what the tarp was used for	0.589	-0.08+	(0.048)	236	H9
Community used the tarp (verified by physical assessment)	0.897	-0.079+	(0.044)	233	H4
Community can show research team the tarp	0.836	-0.116*	(0.051)	232	H5
Respondent would like to be a member of the VDC	0.361	-0.043*	(0.021)	236	H10
Current (or acting) village chief/Headman is younger than 35	0.044	-0.038+	(0.023)	229	H12
Respondent voted in the local government election (2008)	0.851	-0.036*	(0.016)	236	H10
Panel B: Institutions "Improved"					
Community teachers have been trained	0.471	0.122+	(0.066)	173	H4
Respondent is a member of a women's group	0.235	0.060**	(0.021)	236	H8
Someone took minutes at the most recent community meeting	0.295	0.140*	(0.063)	227	H5
Building materials stored in a public place when not in use	0.128	0.246*	(0.098)	84	H5
Chiefdom official did not have the most influence over tarpaulin use	0.543	0.058*	(0.029)	236	H6
Respondent agrees with "Responsible young people can be good leaders" and not "Only older people are mature enough to be leaders"	0.762	0.038*	(0.017)	236	H6, H12
Correctly able to name the Section Chief for this section	0.533	0.053+	(0.032)	234	H9
Correctly able to name the year of the next general elections	0.192	0.038*	(0.018)	236	H9

How should PAP's be used?

- As norms of PAP usage get established, the key question is how much **researcher discretion**. We argue against a "purist" approach with no discretion. Limited flexibility can be desirable but comes with the "price tag" of full transparency: detailed description of all deviations from the PAP; availability of the registered PAP, and complete data sharing.

- We found some degree of flexibility useful. For instance, we deviated from our PAP by adding a 12th hypothesis (on project implementation), to remedy a clear oversight.

Other issues

- Important issues including the timing of PAP registration, the level of analytical detail, and **adjustments for multiple testing**

- **Mean effects** approach (Kling et al 2007) by hypothesis

- **Family wise error rate (FWER)** p-value adjustments to account for multiple testing, for both groups of outcomes and particular outcomes (appendix). Contrast with usual “naïve” or “per comparison” p-values.

Next steps

- The hope is that the registration of PAP's will limit the worst forms of data mining and cherry-picking, and lead to more appropriately sized statistical tests – and as a result boost the credibility of experimental research findings in social science.
- It remains an open question whether norms can be developed that harness these benefits without imposing too great an upfront burden on scholars, or restricting their creativity.

- **Several parallel efforts are currently underway** to establish "trial registries" in the social sciences, including a dedicated committee within the AEA, discussions within the experimental section of APSA, and a planned meeting in Berkeley in December 2012 to forge consensus across disciplines on these issues.

Conclusion

- The project was a reasonable mechanism for delivering local public goods in Sierra Leone, **yet did not lead to lasting changes** in local collective action, village institutions, gender inclusion, social norms.
- The comparative advantage of the World Bank and similar external donors may lie more in **building** development hardware than in **instigating sustainable social change**.

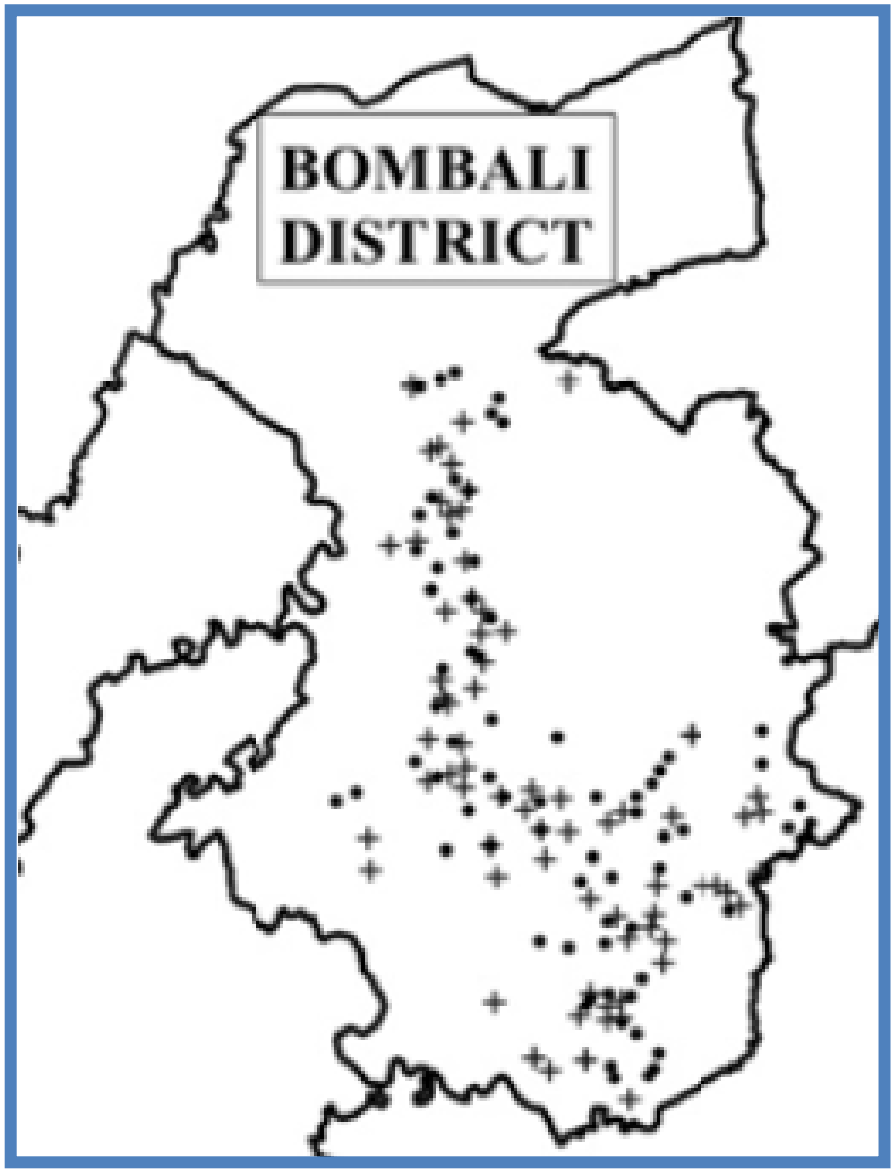
- Setting up new organizations may be insufficient to promote social change since they can be co-opted by elites – here, the chiefs.
- Giving marginalized groups **formal authority** (i.e. Beaman et al 2009 on quotas for women in politics in India) may be more effective than indirect interventions like CDD that hope to shift social norms, especially when existing authorities are strong.

Other evidence on CDD impacts

- **Fearon, Humphreys and Weinstein (2009), Liberia:** No improvement in real-world public goods, material welfare, or meeting attendance in N=83 villages. Higher public goods game contributions in one arm (mixed-gender), plus survey reports of reduced inter-group tension. No funding of small business projects, and no economic impacts.
- **Beath, Christia and Enikolopov (2011), Afghanistan:** Limited impacts on the performance of local institutions and social capital, but some positive impacts on economic well-being, attitudes toward government, and security.
- **Olken (2007), Indonesia:** Top-down audits were more effective in reducing corruption in road projects than grassroots participation.
- **Labonne and Chase (2008), Philippines:** Increased community participation but did not trigger broader social change and may crowd out other activities.
- **Voss (2008), Indonesia:** Mixed impacts on household welfare and access to services: the poor gained, not female headed households.
- Bjorkman and Svensson (2009), Uganda; Banerjee et al. (2010), India.

Robustness checks

- **Were there threats to the research design?**
  - Complete compliance with treatment group assignment
  - Baseline balance on observables across T/C groups
  - Minimal household attrition (4%), moderate for individuals (24%), but balanced across T/C and no interactions with characteristics
- **Did control communities benefit from Gobifo?**
  - GoBifo operated at the ward level as well, so targeting was possible. However, treatment households were, if anything, slightly more likely to report benefits from ward projects (not significant).
- **Are our measures too blunt to detect subtle changes?**
  - Large and diverse number of outcomes for each hypothesis, 318 in all. Consistent results across different data collection methods: HH surveys, direct observation, focus group discussions, and SCAs.


BOMBALI
DISTRICT

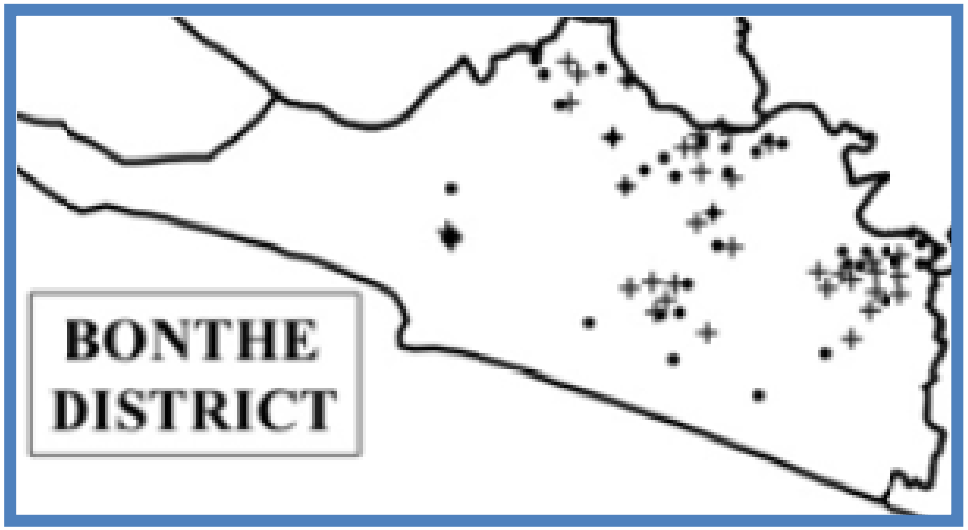

BONTHE
DISTRICT

Table 1: Baseline (2005) Comparison between Treatment and Control Communities

	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)	

Table 2: GoBifo Treatment Effects by Research Hypothesis

Hypotheses by Family	GoBifo Mean Treatment Effect Index	Naïve p-value	FWER adjusted p-value for all 12 hypos	FWER adjusted p-value for 11 hypos in 2009 PAP
	(1)	(2)	(3)	(4)
Family A: Development Infrastructure or "Hardware" Effects				
Mean Effect for Family A (Hypotheses 1 - 3; 39 unique outcomes)	**0.298****			
	(0.031)	**0.000**		
H1: GoBifo project implementation (7 outcomes)	0.703**			
	(0.055)	0.000	0.000	
H2: Participation in GoBifo improves the quality of local public services infrastructure (18 outcomes)	0.204**			
	(0.039)	0.000	0.000	0.000
H3: Participation in GoBifo improves general economic welfare (15 outcomes)	0.376**			
	(0.047)	0.000	0.000	0.000

Table 2: GoBifo Treatment Effects by Research Hypothesis

Hypotheses by Family	GoBifo Mean Treatment Effect Index	Naïve p-value	FWER adjusted p-value for all 12 hypos	FWER adjusted p-value for 11 hypos in 2009 PAP
	(1)	(2)	(3)	(4)
Family B: Institutional and Social Change or "Software" Effects				
Mean Effect for Family B (Hypotheses 4 - 12; 155 unique outcomes)	**0.028 (0.020)**	**0.155**		
H4: Participation in GoBifo increases collective action and contributions to local public goods (15 outcomes)	0.012 (0.037)	0.738	0.980	0.981
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 (47 outcomes)	0.002 (0.032)	0.944	0.980	0.981
H6: GoBifo changes local systems of authority, including the roles and public perception of traditional leaders (chiefs) versus elected local government (25 outcomes)	0.056 (0.037)	0.134	0.664	0.667
H7: Participation in GoBifo increases trust (12 outcomes)	0.042 (0.046)	0.360	0.913	0.914
H8: Participation in GoBifo builds and strengthens community groups and networks (15 outcomes)	0.028 (0.037)	0.450	0.913	0.914

Table 2: GoBifo Treatment Effects by Research Hypothesis

Hypotheses by Family	GoBifo Mean Treatment Effect Index	Naïve p-value	FWER adjusted p-value for all 12 hypos	FWER adjusted p-value for 11 hypos in 2009 PAP
	(1)	(2)	(3)	(4)
H9: Participation in GoBifo increases access to information about local governance (17 outcomes)	0.038 (0.037)	0.301	0.913	0.913
H10: GoBifo increases public participation in local governance (18 outcomes)	0.090* (0.045)	0.045	0.315	0.322
H11: By increasing trust, GoBifo reduces crime and conflict in the community (8 outcomes)	0.010 (0.043)	0.816	0.980	0.981
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 (9 outcomes)	0.041 (0.043)	0.348	0.913	0.914

Table 3: GoBifo Treatment Effects by Hypothesis, Alternative Specifications

Hypotheses by Family	Covariance weighting (Anderson 2008)	SUR approach (Kling and Liebman 2004)	Include panel data	Include full set of controls	Exclude replacement households (attrition)	Include conditional outcomes	Restrict to 2005 hypotheses
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Family A: Development Infrastructure or "Hardware" Effects							
H1: Project Implementation	0.922**	0.700**	0.688**	0.695**	0.706**	0.471**	
	(0.056)	(0.052)	(0.063)	(0.055)	(0.056)	(0.058)	
H2: Local public services	0.233**	0.203**	0.179**	0.206**	0.205**	0.099*	0.149**
	(0.040)	(0.040)	(0.040)	(0.039)	(0.039)	(0.040)	(0.048)
H3: Economic welfare	0.565**	0.371**	0.362**	0.362**	0.375**	0.271**	0.222**
	(0.050)	(0.046)	(0.047)	(0.045)	(0.048)	(0.037)	(0.057)
Family B: Institutional and Social Change or "Software" Effects							
H4: Collective action	-0.043	0.016	0.038	0.011	0.014	-0.040	0.134*
	(0.036)	(0.036)	(0.042)	(0.036)	(0.037)	(0.031)	(0.059)
H5: Inclusion of vulnerable groups	0.000	0.001	0.002	0.000	0.004	0.015	0.067
	(0.029)	(0.030)	(0.030)	(0.031)	(0.032)	(0.027)	(0.116)
H6: Local authority	0.050	0.056	0.051	0.052	0.039	0.053	-0.006
	(0.035)	(0.036)	(0.036)	(0.037)	(0.037)	(0.033)	(0.070)
H7: Trust	0.039	0.042	0.047	0.036	0.048	0.028	0.021
	(0.046)	(0.044)	(0.061)	(0.046)	(0.046)	(0.043)	(0.050)
H8: Groups	0.031	0.027	0.03	0.027	0.045	0.007	-0.048
	(0.037)	(0.035)	(0.039)	(0.037)	(0.037)	(0.034)	(0.054)
H9: Information about governance	0.017	0.037	0.028	0.031	0.045	0.033	0.097*
	(0.038)	(0.035)	(0.040)	(0.036)	(0.037)	(0.035)	(0.043)
H10: Participation in governance	0.160**	0.092**	0.084+	0.082+	0.088+	0.131**	0.088+
	(0.044)	(0.043)	(0.045)	(0.044)	0.046	(0.045)	(0.050)
H11: Crime and conflict	0.041	0.010	0.027	0.014	-0.013	0.011	0.010
	(0.048)	(0.041)	(0.054)	(0.043)	(0.042)	(0.039)	(0.068)
H12: Political and social attitudes	-0.011	0.040	0.040	0.035	-0.011	0.005	
	(0.044)	(0.041)	(0.041)	(0.044)	(0.046)	(0.037)	

Table 4: Illustrative Selection of Statistically Significant Treatment Effects, Family A

Outcome Variable	Mean in Controls	Treatment Effect	Standard Error	N
	(1)	(2)	(3)	(4)
Panel A: Hypothesis 1 - Project Implementation				
Village development committee	0.46	0.40**	(0.05)	235
Visit by WDC member	0.21	0.13*	(0.06)	234
Village development plan	0.62	0.30**	(0.05)	221
Community bank account	0.08	0.71**	(0.05)	226
A local politician was involved in managing the infrastructure:				
Primary School	0.42	0.18**	(0.06)	138
Grain drying floor	0.24	0.13*	(0.06)	115
Latrine	0.22	0.16**	(0.04)	169
Panel B: Hypothesis 2 - Local Public Services				
Functional traditional midwife post in the community	0.08	0.17**	(0.04)	235
Functional latrine in the community	0.46	0.21**	(0.06)	234
Functional community center in the community	0.03	0.09**	(0.03)	236
Community took a proposal to an NGO or donor for funding	0.29	-0.15**	(0.05)	229
Supervisor's physical assessment of construction quality (index from 0 to 1):				
Primary School	0.58	0.11+	(0.06)	123
Grain drying floor	0.38	0.16*	(0.08)	101
Latrine	0.27	0.18**	(0.05)	154
Panel C: Hypothesis 3 - Economic Welfare				
Total petty traders in village	2.43	0.70*	(0.34)	225
Total goods on sale of 10	4.45	0.57*	(0.24)	236
Household asset score	-0.16	0.30**	(0.09)	236
Attended trade skills training	0.06	0.12**	(0.02)	235

Table 5: Illustrative Treatment Effects, Structured Community Activities (SCAs)

Structured Community Activity (SCA) Outcome:	Mean for Controls	Treatment Effect	Standard Error
	(1)	(2)	(3)
Panel A. Collective Action and the Building Materials Vouchers			
GoBifo Mean Effect for SCA #1 (17 outcomes in total)	**0.00**	**0.00**	**(0.05)**
Proportion of communities that redeemed vouchers at building materials store	0.54	-0.02	(0.06)
Average number of vouchers redeemed at the store (out of six)	2.95	0.06	(0.35)
Proportion of communities that held a meeting to discuss the vouchers	0.98	-0.05*	(0.02)
Panel B. Participation in the Gift Choice Deliberation			
GoBifo Mean Effect for SCA #2 (33 outcomes in total)	**0.00**	**0.00**	**(0.04)**
Duration of gift choice deliberation (in minutes)	9.36	1.54	(1.12)
Total adults in attendance at gift choice meeting	54.51	3.57	(2.88)
Total women in attendance at gift choice meeting	24.99	1.98	(1.59)
Total youths (approximately 18-35 years) in attendance at gift choice meeting	23.57	2.06	(1.32)
Total number of public speakers during the deliberation	6.04	0.22	(0.40)
Total number of women who spoke publicly during the deliberation	1.88	-0.20	(0.22)
Total number of youths (approximately 18-35 years) who spoke publicly	2.14	0.23	(0.24)
Proportion of communities that held a vote during the deliberation	0.10	0.07	(0.04)
Panel C. Community Use of the Tarpaulin			
GoBifo Mean Effect for SCA #3 (18 outcomes in total)	**0.00**	**-0.03**	**(0.05)**
Proportion of communities that held a meeting to discuss use of the tarp	0.98	-0.03	(0.02)
Proportion of communities that stored the tarp in a public place	0.06	0.05	(0.04)
Proportion of communities that had used the tarp (5 months after receipt)	0.90	-0.08+	(0.04)
Given tarp used, proportion of communities using the tarp in a public way	0.86	0.02	(0.05)
Proportion of households that directly benefited from the tarp	0.57	-0.01	(0.04)

Appendix G: Sample Attrition by Treatment Group

Dependent variable: Retained in Panel	Individual-level		Household-level	
	(1)	(2)	(3)	(4)
Treatment dummy	-0.017	0.001	-0.011	-0.026
	(0.019)	(0.045)	(0.010)	(0.018)
Treatment * Female		0.012		0.025
		(0.031)		(0.017)
Treatment * Youth (18 to 35 years)		-0.030		0.010
		(0.032)		(0.015)
Treatment * Any education		0.034		0.015
		(0.040)		(0.018)
Treatment * Attended community meeting		-0.018		-0.010
		(0.041)		(0.016)
Treatment * PCA household assets		0.000		-0.005
		(0.012)		(0.007)
Mean retention in panel	0.755	0.755	0.955	0.955
N	2816	2674	2813	2674

Appendix H: Validation of Structured Community Activities (SCAs)

Dependent variable:	Number of vouchers redeemed	Number of women at SCA deliberation	Number of youth at SCA deliberation	Number of women speakers at SCA deliberation	Number of youth speakers at SCA deliberation
	(1)	(2)	(3)	(4)	(5)
Baseline number of functional local public goods (of nine total)	0.263*				
	(0.116)				
Baseline number of female respondents who attended last community meeting		1.289**			
		(0.453)			
Baseline number of youth respondents who attended last community meeting			0.932*		
			(0.403)		
Baseline number of female respondents who spoke at last community meeting				0.159	
				(0.127)	
Baseline number of youth respondents who spoke at last community meeting					0.043
					(0.112)
Controls for total attendance and total speakers					
Total number of attendees at the SCA deliberation		0.437**	0.317**		
		(0.031)	(0.022)		
Baseline number of female respondents surveyed		-0.868		-0.003	
		(0.568)		(0.078)	
Baseline number of youth respondents surveyed			-0.550		-0.028
			(0.376)		(0.066)
Total number of women at the SCA deliberation				0.018+	
				(0.011)	
Total number of youth at the SCA deliberation					0.041**
					(0.011)
N	236	236	236	236	236

Appendix J: "Raw" Results for All Outcomes

Row	Survey question	Hypo-thesis(es)	Outcome type	SCA	Endline mean for controls	Treatment effect	Standard error	Per comparison p-value	FWER p-value (by hypo)	FDR q-value (by hypo)	N
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	Have you personally talked with a member of the WDC or participated in a meeting organized by the WDC in the past year?	H1, H10	full sample		0.090	0.039**	0.013	0.003	0.043; 0.128	0.006; 0.018	236
2	Does this community have a bank account?	H1, H3	full sample		0.081	0.706**	0.045	0.000	0; 0	0.001; 0.001	226
3	In the past year, have you talked with the Local Councillor or participated in any meeting organized by the council?	H1, H10	full sample		0.184	0.028	0.019	0.132	0.531; 0.955	0.059; 0.248	236
4	Since January 2006, has this community had a Village or Community Development Committee (VDC or CDC)?	H1, H4, H10	full sample		0.458	0.399**	0.052	0.000	0; 0; 0	0.001; 0.001; 0.001	235
5	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	full sample		0.617	0.299**	0.048	0.000	0; 0	0.001; 0.001	221
6	Has this community been visited by a Local Council member in the past one year?	H1, H9	full sample		0.322	0.026	0.058	0.653	0.881; 0.997	0.215; 1	236
7	Has this community been visited by a Ward Development Committee member in the past year?	H1, H9	full sample		0.212	0.132*	0.056	0.017	0.151; 0.276	0.018; 0.447	234
8	[Given functional community center in the community] Was a member of the Ward Development committee or Local Council directly involved in the planning, construction, maintenance or oversight of this community center?	H1, H10	conditional		0.238	0.131	0.148	0.288	0.756; 0.995	0.118; 0.326	51
9	[Given functional drying floor in the community] Was a member of the Ward Development committee or Local Council directly involved in the planning, construction, maintenance or oversight of this drying floor?	H1, H10	conditional		0.243	0.128*	0.062	0.029	0.180; 0.563	0.025; 0.087	115

Appendix K: Treatment Effect Heterogeneity Results

	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	0.672**	0.083
	(0.139)	(0.102)
Treatment * Total households in the community	-0.000	-0.001
	(0.001)	(0.001)
Treatment * Index of war Exposure	-0.158	-0.046
	(0.186)	(0.121)
Treatment * Average respondent schooling	-0.018	0.023
	(0.028)	(0.016)
Treatment * Distance to motorable road	-0.006	-0.004
	(0.011)	(0.007)
Treatment * Historical extent of domestic slavery	-0.149*	-0.007
	(0.070)	(0.046)
Treatment * Bombali district	-0.249**	0.033
	(0.063)	(0.045)
Treatment * Ethnolinguistic fractionalization	-0.037	-0.185
	(0.201)	(0.123)
Treatment * Chiefly authority	0.078	0.044
	(0.288)	(0.174)
N	236	236