Preanalysis Plans (PAPs): Applications in Economics

> Katherine Casey Stanford GSB

Summer Institute June 2014



Overview

- Quick scan of pre-registration in Economics
 Early days compared to clinical trials in medicine
- The GoBifo Project: A good fit
 - Design features that posed risks
 - How the PAP mitigated those risks
 - Practicalities in implementing a PAP
- The Elections Project: A tougher fit
 - Ways to build in flexibility when research design demands it
 - Additional upside

Pre-registration in Economics

Clinical Trials in Medicine (what Maya showed you this morning)

RCTs in the new American Economic Association Registry



Source: www.clinicaltrials.gov/ct2/resources/trends

Source: J-PAL Research Newsletter: May 2014

Minimal registration requirements

AEA Trial Registry Required Fields

- Basic identifiers: trial title, country, status, keyword, abstract
- Study timeline:
 - Trial start/end date, intervention start/end date
 - Bonus icon if registered before intervention starts
- Outcomes
- Experimental design (public)
 - Includes planned number of clusters and observations
- IRB approval details (if obtained)

Additional details

- Eligible studies:
 - Open to the social sciences (not just economics), do not need to be an AEA member
 - Observational studies not on the table at the moment
- Options:
 - PAPs not required
 - Can choose to keep different study details (including PAP) private until trial completion
- No current provision for outcome reporting
- Collaborations
 - Developing integrated search to link with OSF, 3ie, EGAP, and other social science registries
 - Will require RCT working papers submitted to NBER to register

Application 1: The GoBifo Project

- Casey, Glennerster and Miguel (2012) estimates the impact of a community driven development program in Sierra Leone on:
 - the "hardware" of local public goods and economic activity, and
 - "software" of institutional performance and social capital
- CDD aims to improve the capacity and performance of local governance while enhancing the inclusion of marginalized groups, like women and youth, in village decision-making
- Research design was a large-scale randomized experiment covering 236 villages over a four year time frame (2005-09) with multiple sources of detailed data collection
- Overall, we found strong positive effects on hardware outcomes and no effects on institutional software

Study features that posed risks

- 1. A vested interest focused on a loosely defined outcome
 - Donors viewed impacts on social capital as a defining feature of CDD
 - Imprecision in what exactly social capital entails provides an "out" that inconvenient results simply capture the wrong measures
- 2. Many relevant outcomes created scope for fruitful cherry picking
 - Institutions are multi-faceted and context-specific
 - Absence of standardized measures makes such tendentious reporting difficult to detect from the outside
- 3. Several sub-groups of theoretical interest
 - *X* sub-groups by *Y* hypotheses invites further cherry picking

How the PAP helped mitigate those risks

- 1. Pre-project (2005) implementation agreement defined the hypothesis set
 - Five hypotheses explicitly capture different dimensions of social capital (trust, collective action, groups, information and inclusion)
- 2. Post-project (2009) PAP defined the universe of outcomes, matched them to specific survey measures, and mapped each one to a hypothesis
 - Provides a credible structure for multiple inference adjustment within and across hypotheses
 - Establishes the hypothesis-level mean effect index as a primary performance metric
 - Commits to providing treatment effect estimates for all 334 outcomes

3. PAP defines 6 primary and 4 secondary sub-groups of interest

- Tests for heterogeneous effects account for multiple inference

PAP disciplines discretion over the raw material...

1 att_wdc	51 smat_well	101 lab_soc	151 nohanghd	201 chf_consult	251 bmon_fish	301 chf_tba
2 bank_acct	52 assets	102 lab_sports	152 noprivol	202 gift_who	252 bmon_seed	302 chf_well
3 met_councilor	53 betteroff	103 lab_tba	153 proj_dem	203 leader_wmn	253 bmon_wom	303 vdc_wmn
4 vdc	54 income	104 lab trad	154 role wmn	204 leader yth	254 bmon youth	304 vdc yth
5 vdp	55 newbiz	105 lab_well	155 role_yth	205 list_lc_chf	255 lab_fish	305 vdp_mat
6 vis_lc	56 out_trader	106 mat_back	156 say_cards	206 not_trad_card	256 lab_seed	306 vdp_tarp
7 vis wdc	57 petty	107 mat comentr	157 say salt	207 not trad salt	257 lab wom	307 vdp writ
8 wdc_comentr	58 quintile	108 mat_dryfir	158 say_tarp	208 not_trad_tarp	258 lab_youth	308 wannabe_VDC
9 wdc_dryfir	59 sold_agric	109 mat_grnstr	159 show_tarp	209 notchf_cards	259 meet_fish	309 no_conflict
10 wdc_grnstr	60 sold_other	110 mat_latrine	160 spkr_tot	210 notchf_salt	260 meet_seed	310 no_fight
11 wdc_latrine	61 tot_goods	111 mat_phu	161 spkr_wmn	211 notchf_tarp	261 meet_wom	311 no_theft
12 wdc_phu	62 tot_newbiz	112 mat_psch	162 spkr_yth	212 notrad_cards	262 meet_youth	312 no_witch
13 wdc_psch	63 tot_petty	113 mat_sports	163 store_tarp	213 notrad_salt	263 ttch_own	313 nobeatchild
14 wdc_sports	64 tot_sources	114 mat_tba	164 tarp_say	214 notrad_tarp	264 disp_ind	314 nobeatwife
15 wdc_tba	65 training	115 mat_well	165 vote	215 question_auth	265 gift_choice	315 nomon_conflict
16 wdc_well	66 ag_income	116 qual	166 wygift_meet	216 resolve_nottrad	266 gift_two	316 violence_bad
17 days brush	67 agric out	117 tchrmoney	167 ben cards	217 rtarp public	267 info gift	317 nomon violence
18 f_barrie	68 market_prod	118 tchrpay	168 goods_show	218 spend_lc_chf	268 info_tarp	318 resolve_within
19 f_comentr	69 other_out	119 train	169 inc_hh	219 tarp_stored	269 name_chr	319 bribebad
20 f dryflr	70 school days	120 used cards	170 meet farm	220 trust lc chf	270 name elec	320 noforcework
21 f_gmstr	71 anycard	121 ben_salt	171 meet_lab	221 tstore_notchf	271 name_lc	321 strangeok
22 f_latrine	72 brush	122 ben_tarp	172 meet_osu	222 mstore_pub	272 name_pc	322 vh_fem
23 f_market	73 cards	123 debate	173 meet_pta	223 send_not_trad	273 name_proj	323 vh_youth
24 f_palava	74 commfarm	124 dem_cards	174 meet_rel	224 tchr_dec	274 name_sc	324 youthtreat
25 f_phu	75 commtchr	125 dem_salt	175 meet_sav	225 tchr_rep	275 name_tax	325 frm_age
26 f_psch	76 mkt_grp	126 dem_tarp	176 meet_soc	226 hmarket	276 radio	326 frm_nokid
27 f_well	77 tarp	127 democ	177 meet_tchr	227 hwallet	277 vis_pc	327 frm_sex
28 footunif	78 tarp_freq	128 disabled_ldr	178 meet_trad	228 osusu	278 info_cards	328 frm_trb
29 func_sports	79 vchr_self	129 disabled_meet	179 mtng_comentr	229 rmarket	279 name_dues	329 groupsave_ind
30 func tba	80 vchr tot	130 duration	180 mtng dryfir	230 trust og	280 change chiefdom	330 labgang ind
31 no_bush	81 wkcomfrm	131 equal_cards	181 mtng_grnstr	231 trust_chf	281 change_council	331 osusu_ind
32 proposal	82 bmon_lab	132 equal_salt	182 mtng_latrine	232 trust_lc	282 council_listen	332 religgroup_ind
33 seedbank	83 bmon osu	133 equal tarp	183 mtng phu	233 trust ngo	283 cvote local	333 sociald ind
34 tarp_public	84 bmon_pta	134 gift_big	184 mtng_psch	234 trust_out	284 cvote_pres1	334 tradsoc_ind
35 card_public	85 bmon_rel	135 gift_dem	185 mtng_sports	235 trust_own	285 cvote_pres2	
36 cf_barrie	86 bmon_sav	136 gift_meet	186 mtng_tba	236 trust_pol	286 discuss_politics	
37 cf_comentr	87 bmon_soc	137 gift_say	187 mtng_well	237 rwallet	287 stand_lc	
38 cf_dryfir	88 bmon_trad	138 maj_gift	188 pwy_hh	238 chumos	288 stand_pc	
39 cf_latrine	89 daysfrm	139 meet_cards	189 rcpt_cards	239 dues	289 stand_sc	
40 cf psch	90 lab comentr	140 meet com	190 recd cards	240 fishcoop	290 stand wdc	
41 cf_well	91 lab_dryfir	141 meet_salt	191 show_mat	241 mbr_fish	291 vote_local	
42 fin_sports	92 lab_grnstr	142 meet_tarp	192 spk_cards	242 mbr_pta	292 vote_pres1	
43 fin tba	93 lab lab	143 meet tot	193 spk com	243 mbr rel	293 vote pres2	
44 qual_dry	94 lab_latrine	144 meet_wmn	194 spk_farm	244 mbr_sav	294 chf_comentr	
45 qual_lat	95 lab_osu	145 meet_yth	195 spk_gift	245 mbr_seed	295 chf_dryflr	
46 qual_psch	96 lab_phu	146 minutes	196 spk_salt	246 mbr_soc	296 chf_grnstr	
4/ qual_well	9/ lab_psch	14/ mtng_cards	197 spk_tarp	24/ mbr_trad	297 chf_latrine	
48 smat_dry	98 lab_pta	148 mtng_salt	198 spk_tchr	248 mbr_wom	298 chf_phu	
49 smat_lat	99 lab_rel	149 mtng_tarp	199 store_mat	249 mbr_youth	299 chf_psch	
50 smat_psch	100 lab_sav	150 nocorrupt	200 wide_pay	250 ttch_oth	300 chf_sports	

June 2014

Into a clear set of results with high internal validity

TABLE II

GoBifo Treatment Effects by Research Hypothesis						
Hypotheses by family	(1) GoBifo mean treatment effect endex	(2) Naive <i>p</i> -value	(3) FWER-adjusted <i>p</i> -value for all 12 hypos	(4) FWER-adjusted <i>p</i> -value for 11 hypos in 2009 PAP		
Family A: Development infrastructure or "hardware" effects						
Mean effect for family A (Hypotheses 1-3; 39 unique	0.298**					
outcomes)	(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	0.204^{**}					
public services infrastructure (18 outcomes)	(0.039)	0.000	0.000	0.000		
H3: Participation in GoBifo improves general economic wel	0.376^{**}					
fare (15 outcomes)	(0.047)	0.000	0.000	0.000		
Family B: Institutional and social change or "software" effects						
Mean effect for family B (Hypotheses 4–12; 155 unique	0.028					
outcomes)	(0.020)	0.155				
H4: Participation in GoBifo increases collective action and	0.012					
contributions to local public goods (15 outcomes)	(0.037)	0.738	0.980	0.981		
H5: GoBifo increases inclusion and participation in	0.002					
community planning and implementation, especially for	(0.032)	0.944	0.980	0.981		
poor and vulnerable groups; GoBifo norms spill over into						
other types of community decisions, making them more						
Inclusive, transparent, and accountable (47 outcomes)	0.050					
Ho: Gobilo changes local systems of authority, including the	0.056	0.194	0.004	0.007		
roles and public perception of traditional leaders (chiefs)	(0.037)	0.134	0.664	0.667		
versus elected local government (25 outcomes)						

How does this work?

- PAP document specifies:
 - Hypotheses and outcomes
 - Distinguish primary from secondary outcomes if relevant
 - Link outcomes to specific survey measures
 - Group outcomes into hypotheses / families
 - Econometric specifications
 - Design basics
 - Control set
 - Stratification variables
 - Clustering level, observations per cluster
 - Dimensions of heterogeneous treatment effects / sub-group analysis
 - Mean effects by level if relevant

Timeline

Appendix B: Project and Research Timeline

10-Oct-05	\downarrow	Hypothesis document drafted	Jan-08		
Nov-05		Baseline Survey	Feb-08		Projects implemented
Dec-05	\downarrow	Dasenne Sarvey	Mar-08	\downarrow	
Jan-06			Apr-08		Second grants disbursed
Feb-06		Ward Facilitator Training	May-08	↓.	
Mar-06		Ward Faelinator Franking	Jun-08		
Apr-06	\downarrow		Jul-08		Projects implemented
May-06			Aug-08	↓.	
Jun-06			Sep-08		Third grants disbursed
Jul-06			Oct-08	\downarrow	Third grants disoursed
Aug-06		Development Planning	Nov-08		
Sep-06		Development Framming	Dec-08		
Oct-06	5		Jan-09		Projects implemented
Nov-06			Feb-09		r tojects implemented
Dec-06	\checkmark		Mar-09		
Jan-07		Ward Davalonment Committee	Apr-09	\downarrow	
Feb-07		Approval	May-09	\downarrow	Follow-up survey 1
Mar-07	\mathbf{V}	Аррюча	Jun-09		Voucher program begins
Apr-07			Jul-09	\checkmark	votiener program begins
May-07			21-Aug-09	1	Pre-Analysis Plan archived with the
Jun-07				\checkmark	Jameel Poverty Action Lab
Jul-07		Dalaur	Sep-09	↓	Voucher program ends
Aug-07	T	Delays	Oct-09		E-11
Sep-07			Nov-09	\downarrow	Follow-up survey 2
Nov-07			4-Mar-10	1	Plan Supplement covering second
Dec-07	Ý.			Ý	follow-up survey archived

What the GoBifo PAP looks like

Community Driven Development in Sierra Leone: GoBifo Analysis Plan

Final version: August 21, 2009

PIs: Rachel Glennerster Edward Miguel

This document outlines the plan for analyzing the impact of the GoBifo Project, using the endline round 1 data. Note that this document was written up before the analysis of any endline round 1 data. We will produce a similar document before the analysis of any GoBifo endline round 2 data, which has not yet been collected.

Table of Contents:

I. Overview II. Regression Specifications III. Hypotheses:

H1: Participation in GoBifo increases trust

H2: Participation in GoBifo increases collective action and contribution to local public goods.

H3: Participation in GoBifo improves the quality of local public services infrastructure.

The working document

Final version: August 21, 2009 PIs: Rachel Glennerster Edward Miguel

This document outlines the plan for analyzing the impact of the GoBifo Project, using the endline round 1 data. Note that this document was written up before the analysis of any endline round 1 data. We will produce a similar document before the analysis of any GoBifo endline round 2 data, which has not yet been collected.

Table of Contents:



Econometric specifications

II. Regression specifications

II.A. General Framework

The most general strategy for testing each hypothesis will be to regress the measures relevant for each hypothesis on a treatment indicator variable and controls using the following model:

 $Y_{ic} = \beta_0 + \beta_1 T_c + V_c^{'} \Gamma + W_c^{'} \Pi + \mathcal{E}_{ic}$

where Y_{ic} is a given outcome (e.g., participation in local road brushing activities) for household *i* in community *c*; T_c is the village treatment dummy; V_c is a vector of the community level controls; W_c is a fixed effect for geographic ward, the administrative level on which the randomization was stratified; and ε_{ic} is the usual idiosyncratic error term, clustered at the village level (the unit of randomization). Here the parameter of interest is β_1 , the average treatment effect. Note that V_c can either be a sparse set of community level controls such as distance from road, population size, or a more detailed set of controls, including all the variables for which we expect interaction effects, as discussed below in section. The analysis will present specifications with both the sparse and detailed V, as each have their possible strengths, e.g., while both yield unbiased estimates of program impacts, the more saturated specification may benefit from more

precise estimates (smaller standard errors).

For all outcomes that were collected in both the baseline and endline surveys, analysis will exploit the panel structure of the data using the following adapted model:

$$Y_{ict} = \beta_0 + \beta_1 T_c + \beta_2 P_t + \beta_3 (T_c \times P_t) + V'_c \Gamma + W'_c \Pi + \varepsilon_{ict}$$

where Y_{ict} is a particular outcome for household *i* in community *c* at time *t*, where t = 0 if the June 20 observation was recorded before the program began (in the baseline survey) and t = 1 if recorded

Econometric specifications (cont.)

The discussion of hypotheses below lists each indicator from the baseline and/or endline surveys that will be used to test each hypothesis. Standard errors in regressions using household level data will be adjusted to account for the fact that treatment is at the village level, by clustering disturbance terms by village. For each hypothesis, Y_{ic} (or Y_c) will be evaluated at least two separate ways:

1) regressing a single outcome measure on the dependent variables specified above; and 2) "mean effects" estimation, using multiple outcome measures to evaluate if the program has had an impact on a set of closely inter-related outcomes, for instance, the multiple questions dealing with trust, or those measuring information about local governance and politics, or local public service infrastructure, among others (as in Kling et al. 2007).

Table III: Sensitivity to specification choices

TABLE III GoBifo Treatment Effects by Hypothesis, Alternative Specifications								
002	GODIFO TREATMENT EFFECTS BI TITFOTRESS, ALTERNATIVE SPECIFICATIONS							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
	Covariance	SUR	Include	Include full	Exclude	Include	Restrict	
	weighting	approach	panel	set of	replacement	conditional	to 2005	
	(Anderson 2008)	(Kling and	data	controls	households	outcomes	hypotheses	
Hypotheses by family		Liebman 2004)			(attrition)			
Family A: Development infrastructur	e or "hardware" ef	fects						
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)	

Notes: Significance levels (naive p-value) indicated by p < .10, p < .05, p < .01. Robust standard errors in parentheses. Includes fixed effects for the district council wards (the unit of stratification) and the two balancing variables from the original randomization—total households per community and distance to nearest motorable road. Outcomes included per hypothesis vary by column: columns (1)–(5) include full sample outcomes only (184 unique outcomes in total), column (6) includes both full sample and conditional outcomes (i.e., those that depend on the state of another variable, e.g., quality of infrastructure depends on the existence of the infrastructure, 334 unique outcomes in total), and column (7) includes 63 unique outcomes. Column (1) weights each index component by the inverse of the appropriate element of the variance-covariance matrix (as in Anderson 2008) where the matrix is estimated in the control group (zero replaces any negative estimated weights). Column (2) uses stacked ordinary least squares outcome-by-outcome as in Kling and Liebman (2004). Column (3) uses the Kling and Liebman (2004) approach incorporating panel data where available. Column (4) uses the Kling, Liebman, and Katz (2007) approach into the full set of outcomes (note five variables from the PAP were omitted due to insufficient observations: community financial contributions to peripheral health unit, palava hut, market, and grain store (H2 and H4) and existence of football equipment (H2)). Column (7) uses Kling. Liebman, and Katz (2007) restricted to the hypothese written down in the 2005 program downeet and the full sample outcomes included to the hypothese written down in the 2005 program downeet and to full sample outcomes included in the baseline of the hypothese written down in the 2005 program downeet and to full sample outcomes included in the hypothese written down in the 2005 program.

2005 survey.

Econometric specifications (cont. 2)

II.B. Interaction Effects

- Household socioeconomic status (e.g., education, asset ownership)ⁱⁱ
 - Similar to the hypotheses for women and youth, poorer households were targeted by the program for greater voice in local community governance and thus may benefit more than other households. However, their marginalized position may have prevented them from capturing GoBifo benefits relative to other households.
- District (Bombali vs. Bonthe)
 - Randomization was stratified by district, and program effects may plausibly differ across districts due to their different ethno-linguistic, socio-economic and institutional characteristics, issues that we intend explore in detail.
- Indicators of remoteness (e.g. distance to roads).
 - At baseline, remote communities may be poorer, have less information, and less access to government officials and NGOs than less remote communities. They may also be more cohesive with less in and out migration or community members working outside the community. The value of materials communities could purchase with fixed GoBifo grants was less given the very high transport costs incurred in bringing the materials to the communities (a concern raised by GoBifo staff). For these reasons we might expect differential program impacts in more remote areas.
 - Community size
 - In our discussions with GoBifo field staff, many indicate that they believe

where *R_i* we hypot available

We are in

villages v

this end.

indicator

 $Y_{ict} = \beta_0$

DITES SUMMER MISULULE

Heterogeneous effects appendix table

	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

Outcomes by hypothesis

H3: Participation in GoBifo improves the quality and quantity of local public services infrastructure.

Community Level outcomes: Primary (all panel data) Typo: should read "grain store" not "drying floor."

- Treatment communities have more/higher quality primary schools than controls (Village module, C1B and C1C; K10A through K10D).
- Given that the community has a primary school, a higher share of treatment communities provide community funds to it (completely or partially) (Village module, C1D)
- Treatment communities have more/higher quality public health units (community health centers, community health posts, maternal & child health post) than controls (Village module, C3B, C3C, C3AB).
- Given that the community has a public health units (community health centers, community health posts, maternal & child health post), a higher share of treatment communities provide community funds to it (completely or partially) (Village module, C3D)
- Treatment communities have more/higher quality water wells (manual or mechanical wells) than controls (Village module, C4B, C4AB, C4BB; K13A through K13D).
- Given that the community has a well, a higher share of treatment communities provide community funds to it (completely or partially) (Village module, C4AC, C4BC).
- Treatment communities have more/higher quality drying floors than controls (Village module, C7B and C7C).
- Given that the community has drying floors, a higher share of treatment communities provide community funds to it (completely or partially) (Village module, C7D).
- Treatment communities have more/higher quality communal grain stores than controls (Village module, C8B and C8C; K12A through K12D^{xii}).
- Given that the community has drying floors^{xiii}, a higher share of treatment

Primary results table

TABLE II GoBifo Treatment Effects by Research Hypothesis

Hypotheses by family	(1) GoBifo mean treatment effect endex	(2) Naive <i>p</i> -value	(3) FWER-adjusted <i>p</i> -value for all 12 hypos	(4) FWER-adjusted <i>p</i> -value for 11 hypos in 2009 PAP
 Family A: Development infrastructure or "hardware" effects Mean effect for family A (Hypotheses 1-3; 39 unique outcomes) H1: GoBifo project implementation (7 outcomes) 	0.298** (0.031) 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 GoBilo improves general economic wel fare (15 outcomes) Family B: Institutional and social change or "software" effects	(0.047)	0.000	0.000	0.000
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: GoBio 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)	(0.032)	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

"Raw results" appendix table

Row	v 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)
20	Does the community have a drying floor and is it functional?	H2	full sample		0.237	0.160**	0.055	0.004	0.11	0.015	228
21	Does the community have a grain store and is it	H2	full sample		0.136	0.067	0.045	0.135	0.907	0.156	235
22	Does the community have a latrine and is it functional?	H2	full sample		0.462	0.208**	0.059	0.001	0.019	0.005	234
23	Does the community have a market and is it functional?	H2	full sample		0.017	-0.001	0.016	0.976	1	0.641	235
24	Does the community have a palava hut and is it functional?	H2	full sample		0.096	-0.004	0.037	0.923	1	0.634	231
25	Does the community have a public health unit and is it functional?	H2	full sample		0.060	0.017	0.032	0.595	1	0.523	235
26	Does the community have a primary school and is it functional?	H2	full sample		0.462	0.071	0.057	0.206	0.963	0.209	234
27	Does the community have any wells (mechanical or bucket) and are any of them functional?	H2	full sample		0.459	0.032	0.063	0.604	1	0.523	222
28	Do any of the local sports teams have uniforms / vests?	H2	full sample		0.100	0.102*	0.048	0.031	0.512	0.068	225
29	Does the community have a football / sports field and is it functional?	H2	full sample		0.444	0.069+	0.040	0.089	0.813	0.128	236
30	Does the community have a traditional birth attendant (TBA) house and is it functional?	H2	full sample		0.079	0.172**	0.035	0.000	0	0.001	235
31	Ask to be taken to the nearest bush path. This should be a foot path (not a road for cars) that the community uses the most. Walk 100 steps down the path (i.e. look at the middle, not the start of the path). In your own opinion, how bushy is the path? [Answer indexed from 0 "very bushy" to 1 "very clear"]	H2, H4	full sample		0.482	-0.003	0.034	0.942	1; 1	0.634; 1	228
45	Supervisor summary assessment of the overall appearance of the latrine (index from $1 =$ excellent to $0 =$ unfit for use) June 2014	H2	conditional BITSS S	Summe	0.417 er Institute	0.060+	0.031	0.047	0.644	0.087	153

Why this matters: The paper we could have written

TABLE VI Erroneous Interpretations under "Cherry Picking"

Outcome variable	(1) Mean for controls	(2) Treatment effect
Panel A: GoBifo "weakened" institutions		
Attended meeting to decide what to do with the tarp	0.81	-0.04^{+}
Everybody had equal say in deciding how to use the tarp	0.51	-0.11^{+}
Community used the tarp (verified by physical assessment)	0.90	-0.08^{+}
Community can show research team the tarp	0.84	-0.12*
Respondent would like to be a member of the VDC	0.36	-0.04*
Respondent voted in the local government election (2008)	0.85	-0.04^{*}

Why this matters: The paper we could have written (v2)

TABLE VI

ERRONEOUS INTERPRETATIONS UNDER "CHERRY PICKING"

	(1)	(2)
	Mean for	Treatment
Outcome variable	controls	effect

Panel B: GoBifo "strengthened" institutions		
Community teachers have been trained	0.47	0.12^+
Respondent is a member of a women's group	0.24	0.06**
Someone took minutes at the most recent community meeting	0.30	0.14*
Building materials stored in a public place when not in use	0.13	0.25*
Chiefdom official did not have the most influence over tarp use	0.54	0.06*
Respondent agrees with "Responsible young people can be good leaders" and not "Only older people are mature enough to be leaders"	0.76	0.04*
Correctly able to name the year of the next general elections	0.19	0.04*

Incorporating omissions and learning

- We forgot things: added a hypothesis ex post regarding project implementation by drawing together outcomes already in the PAP
- We learned from research fieldwork and piloting: developed new measures of collective action (e.g. SCAs); threw out baseline measures with little variance
- We acquired new information from program implementation: did not anticipate the focus on skills training, so added new measures to the endline survey
- We added framing to ease interpretation: grouped hypotheses under two intuitive families ex post

A compromise: Limited flexibility with full transparency

- Some flexibility is useful to counter downside risks of a "purist" approach
 - Rigidity may still learning or limit leverage of all available information
 - Excessive up front costs may deter adoption
- ... If it is accompanied by transparency to maintain the credibility of the pre-specification process
 - Report results with and without ex post adjustment
 - Identify what was pre-specified and when to allow readers to make their own informed judgments

Application 2: The Elections Project

- Bidwell, Casey and Glennerster (ongoing) study the impact of debates between Parliamentary candidates on voter knowledge, policy alignment and vote choice, as well as candidate spending and politician performance
- Key differences from the GoBifo application
 - Very tight implementation timeline: a matter of weeks between official announcement of candidates and Election Day
 - Early implementation/data collection stages designed to inform later stages, but not enough time to process and analyze data in between (pre-specification useful for planning, survey writing)
 - Cherry picking less of a risk as primary data source is a 15 minute exit poll with relatively few outcomes
 - Built more "upside" into the PAP

Timeline



How we built in some flexibility

- From a purist perspective, we specified the main PAP governing the final stage exit poll outcomes first, while the exit poll was still in the field
- To accommodate flexibility, that first PAP lays out the planned series of intermediate analyses including how earlier stages would inform later stages
- We lodged a separate PAP for the intermediate stages before looking at that earlier data
- After conducting the intermediate analysis, we lodged a revision to the main PAP before analyzing the final data

1st PAP governs ultimate final stage outcomes

SIERRA LEONE 2012 ELECTIONS PROJECT

PRE-ANALYSIS PLAN: POLLING CENTER LEVEL INTERVENTIONS

PIs: Kelly Bidwell (IPA), Katherine Casey (Stanford GSB) and Rachel Glennerster (JPAL MIT)

20 November 2012

This study examines the impact of providing citizens with information about Parliamentary candidates via structured inter-party debates in the lead up to the Sierra Leone November 2012 Elections. Randomization and treatments were conducted on multiple levels: constituency, polling center and individual (details on sampling and randomization are available in the project's "Sampling Procedures" document). This pre-analysis plan governs the analysis of the polling-center level treatment only. It was written and registered with the Abdul Latif Jameel Poverty Action Lab before fieldwork for the exit poll, which is the primary source of data for this analysis, was completed (where the current estimated completion date is 22 November 2012). This document is the first installment in a planned sequence of registry and data analysis, where we will next: (i) register separate plans for the individual-level and constituency-level treatments; (ii) analyze treatment effects for the individual-level treatments; (iii) analyze the expert panel scoring of debates and the before/after debate surveys; (v) register an update to this document reflecting learning from steps 2 to 4; and then (vi) analyze treatment effects at the polling-center level in the exit poll and voting returns data

Separate PAPs for intermediate stage

SIERRA LEONE 2012 ELECTIONS PROJECT

PRE-ANALYSIS PLAN: INDIVIDUAL LEVEL INTERVENTIONS

PIs: Kelly Bidwell (IPA), Katherine Casey (Stanford GSB) and Rachel Glennerster (JPAL MIT)

THIS DRAFT: 15 August 2013

This study examines the impact of providing citizens with information about Parliamentary candidates via structured inter-party debates in the lead up to the Sierra Leone November 2012 Elections. Randomization and treatments were conducted on multiple levels: constituency, polling center and individual (details on sampling and randomization are available in the project's AEA trial registry). This pre-analysis plan governs the analysis of the individual level treatments only. It was written and registered before analysis of the individual treatments data. It incorporates learning from analysis of the before/after screening data within the PC-level treatment sites.

Revised Final Stage PAP

SIERRA LEONE 2012 ELECTIONS PROJECT

PRE-ANALYSIS PLAN: POLLING CENTER LEVEL INTERVENTIONS

PIs: Kelly Bidwell (IPA), Katherine Casey (Stanford GSB) and Rachel Glennerster (JPAL MIT)

Revised Plan: 12 September 2013

This study examines the impact of providing citizens with information about Parliamentary candidates via structured inter-party debates in the lead up to the Sierra Leone November 2012 Elections. Randomization and treatments were conducted on multiple levels: constituency, polling center and individual (details on sampling and randomization are available in the project's AEA trial registry https://www.socialscienceregistry.org/trials/26). This pre-analysis plan governs the analysis of the polling-center level treatment only. The first version of this plan was written and registered with the Abdul Latif Jameel Poverty Action Lab on 20 November 2012, before fieldwork for the exit poll, which is the primary source of data for this analysis, was completed. This revised plan incorporates learning from the following steps that we have taken since registering the initial plan, namely we: (ii) analyzed the expert panel scoring of debates and the before/after debate surveys; (ii) registered a separate plan for the individual-level treatments; (iii) analyzed treatment effects for the individual-level treatments; and (iv) examined the distribution of outcomes for the control group polling centers in the exit poll data. We are now registering an update to the initial document reflecting learning from steps 1 to 4; before we analyze treatment effects at the polling-center level in the exit poll. Planned future steps include: i) lodging an update governing the analysis of the electoral returns data before completing that portion of the analysis (which depends on two additional datasets that have not yet been cleaned); and ii) lodging an update governing the analysis of constituency-level effects (as this data collection effort remains ongoing),

Comment [KC1]: For transparency, we have tracked the changes we made to the original PClevel PAP lodged on 20 Nov 2012 and included explanatory comments for the more substantive revisions.

Deleted: 20 November 2012

Deleted: "Sampling Procedures" document

Deleted: It

Comment [KC2]: We changed the planned ord of our analysis to complete more of the explorator work before embarking on the PC-level analysis.

Deleted: (where the current estimated completion date is 22 November 2012). This document is the first installment in a planned sequence of registry and data analysis, where we will next:

Deleted: s

Deleted: and constituency-level

Deleted: ii

Deleted: ; (iv) analyze the expert panel scoring of

Learning and algorithms to choose controls

- We specified how we would choose control variables after looking at the data
 - In 1st PAP:

center); W is a set of additional control variables that will be determined from analysis of the control group data and will vary by hypothesis with an eye toward identifying individual characteristics that do not vary with treatment and that help explain variation in a particular outcome (i.e. education and radio ownership are likely positively correlated with general political knowledge); c is a set of constituency-

- In Revised PAP:

center); *W* is a set of additional control variables determined from analysis of the control group data and will vary by hypothesis with an eye toward identifying individual characteristics that do not vary with treatment and that help explain variation in a particular outcome (see algorithm below); *c* is a set of constituency-specific fixed effects (the level of debate and candidates); and ε is an idiosyncratic error term clustered at the polling center level. Our main specification includes the full set of controls (*X*, *Z* and *W*); we will also show results for the sparser specification that includes only the stratification variables as controls (*X* and *Z* only) as a robustness check. We will determine *W* as any subset of {gender, age, frequency of discussing politics, education, marital status, occupation, radio ownership} that predicts outcomes for the control group with at least 95% confidence. The coefficient of interest is δ ,

Upside: One-sided tests

• For outcomes with a clear theoretically predicted direction, we pre-specified one-sided tests

Vote choice outcomes

- Tests to conduct $\delta_t \ge 0$ for $t \in D, R, G$; $\delta_{t \in D, R, G} \ge 0$; $\delta_t \ne \delta_{\sim t}$ for $t \in D, R, G$
- a. Hypothesis 1: Exposure to debates increases vote shares for the candidate that performed the best in the debates
 - i. TE measured by vote choice
 - ii. Debate winner / loser measured by audience ratings and expert assessment
- b. Hypothesis 2: Exposure to debates increases the willingness to vote across party lines
 - i. TE measured by vote choice and ethnicity (limited to members of affiliated tribes)
 - ii. Three additional (also primary) measures expand the concept to incorporate nonaffiliated tribes: voting for a different party for MP in 2012 than in 2007; splitting ticket for MP (i.e. party MP different than party Pres or party LC); and voting for a party other than the self-reported party affiliation from before survey.

Upside: Bolstering descriptive analysis

• Pre-specified potential causal mechanisms to add credibility to eventual descriptive analysis and inference

Mechanism of impact

- i. *Comprehension and attention* may vary by mode of information delivery. A finding that $\delta_D > \delta_R$ for general political knowledge questions (H3) suggests that debates may better engage the audience than radio summaries. Check for waning attention by placement of knowledge questions in the program (i.e. MP roles at the beginning, date of election at the end)
- For D, the impact on correctly locating candidate positions should be increasing in the performance of the candidates in answering policy questions as assessed by the expert panel.

Conclusion

- Pre-analysis plans (PAPs) help enhance the credibility of research
- Pre-specification and PAPs are still in very early stages in economics
- As norms evolve, one strategy to accommodate learning is limited flexibility with complete transparency
- Include the most stringent "purist" specifications as a benchmark for more flexible or *ex post* adjustments
- PAPs are not without costs, but offer opportunities for upside as well