WORKING DRAFT - PLEASE DO NOT CITE

Empowering Citizens to Combat the Resource Curse

DESIGN MEMO

Laura Paler
Columbia University
April 28, 2010

1 Overview

This memo describes sampling and randomization for the research project "Empowering Citizens to Combat the Resource Curse." The project has two components: (1) a field experiment consisting of a randomized information campaign with four treatment conditions; and (2) a data collection effort, including a survey and a postcard campaign (a behavioral measure). (For more information on the theoretical motivation for this project and the design of the treatments, see the accompanying Theory Memo.) From March-May 2010, the project was implemented in 93 randomly sampled villages around Blora achieving a total target sample of 1862 individuals. The project was conducted in partnership with the Indonesian NGOs PATTIRO and LPAW, who were consulted extensively during the design stages to ensure a realistic implementation plan given local context.

Individual respondents were identified using multi-stage cluster sampling, with 20 individuals selected per village. Assignment to different treatment conditions was done at the individual level and blocked at the village level, with five individuals per treatment condition per village (an average of 465 per treatment condition overall). This memo describes the power for detecting treatment effects, sampling, treatment assignment, and data analysis plans.

2 Target Population

The population of interest for the survey are all adults living in Blora. For survey purposes, this was defined as anyone from 17-65 years of age who has lived in Blora for at least the previous six months. The target populations corresponds to the population of eligible voters; in Indonesia, anyone over 17 or who is already married is eligible to vote. Based on the 2008 PODES survey, which is conducted every 2-3 years in every village, the total population in Blora in 2007 was estimated to be 904,637 people in 247,381 households. While there is no exact estimate of the target population, the number of voters registered for the 2009 presidential elections–686,635 individuals–can serve as a lower bound. Updated population information was collected during project implementation and will be used to construct the sample weights.

3 Power Analysis

The goal of the project was to test hypotheses on the impact of tax payments and transparency on demand for good government. To test hypotheses we used a randomized information campaign with independent tax and transparency treatments. The experiments were implemented according to a 2x2 factorial design, producing four treatment conditions (see Table 1). The treatments were also designed in a way to enable direct comparisons across all treatment conditions. Consequently, we wanted to ensure we had a sample of sufficient size to detect treatment effects across the six combinations of comparisons.

Figure 1: Treatments

		Tax Experiment	
		Control (***Pure Windfall, Taxes=0***)	**Treatment** (***Taxes=4,000 rupiah***)
Transparency Treatment	**Control** ***(Placebo information)***	**1** n=465 n per village=5 Number of villages=93	**2** n=465 n per village=5 Number of villages=93
	Treatment ***(Spending information)***	**3** n=465 n per village=5 Number of villages=93	**4** n=465 n per village=5 Number of villages=93

The individual was the unit of randomization. While randomized information campaigns are often done at the cluster level (Gerber and Green, 2000; Arceneaux, 2005), there were several reasons to focus on the individual level for this study. First, the tax experiment involved providing respondents with income and deducting a representative tax to the district government, which would have been difficult to implement at a group level. Second, for the transparency treatment, we were interested in estimating the impact of information on all adults in Blora, including those who typically do not seek out such information. By targeting randomly sampled individuals with the campaign, we increased the likelihood that these individuals would be reached.

Additionally, treatment assignment was blocked at the village/dusun level, meaning that individuals were assigned to treatment groups within villages. Blocking increases power by ensuring that treatment and control units are closely matched within groups, reducing within site heterogeneity. Blocking at the village level can be modeled as a two level hierarchical linear model, where the first level is the individual and the second level is the group (village/dusun). The implication for analysis is that any linear model should adjust standard errors to account for group-level clustering (see Section 7.

I used Optimal Design 2.0 software for power analysis for an individual level treatment assignment with randomization blocked at the village/dusun level. Using their notation, let n be the number of persons per site, J be the number of sites, B be the percent of variance explained by the blocking variable (village), δ be the (standardized) treatment effect size, and σ^2_δ be the variance of the standardized treatment effect.[1] Rather than select n and J to achieve a specific minimal detectable effect size (MDES), n and J were determined first for several practical reasons. First, n and J were determined primarily by time and budget constraints. Second, I had little ex ante expectation of what δ and σ^2_δ would be. While some survey data is available, nothing matched the overall project design close enough to provide a reliable basis for analysis. I estimated that one team could complete three villages per week and each campaigner could complete two visits per day. Under our budget constraint, we could field three teams with five campaigners per team, producing a total J of 93 and a total n of 1860.

Figure 2 shows for power=.80 the minimum effect size that can be detected for different values of the variance for these combinations of J and n. Power was greater with J=96 than with J=93, which is consistent with a large literature that shows power increases more with J than with n. For practical implementation purposes we selected J=93 and n=20, producing a total sample of 1860, with 465 individuals per treatment condition.

Taking J=93 and n=10, I next looked at the minimum detectable effect size that I could obtain as a function of σ^2_δ, given different assumptions about power and B. Figure 3 plots the standardized effect size on the y-axis and treatment effect variability on the x-axis for power=.80 and .90 and for B=0 and B=.20. This approach helps to see the gains from blocking as B increases. It is also useful because once I have the real data I will be able to estimate σ^2_δ. I will then be able to back out what effect size, given σ^2_δ, I should be able to detect, allowing me to judge whether insignificant results are due to insufficient power or no treatment effect.

4 Sampling

Sampling for this project follows a multi-stage cluster design where villages are the primary sampling unit, followed by dusun, households and main respondents (all selected using simple random sampling). All sampling protocols can be found in the Field Manual.

4.1 Strata and Primary Sampling Units

The administrative hierarchy in Indonesia is: District (*kabupaten)* → subdistricts (*kecamatan)* → villages (*desas/kelurahan*) → dusun/RW → RT (*rukun tetangga*). Villages served as the primary sampling unit (PSU). There are 295 villages in Blora in 16 subdistricts. To obtain the target of 93 villages, strata were first created by subdistrict and urban/rural designation. A village was deemed urban if it was officially classified as a *kelurahan* (rather than a *desa*) according to 2008 BPS Blora data. There are a total of 25 kelurahan in Blora, primarily clustered in the urban centers of Blora and Cepu. Proportional random sampling was used to select PSUs. Within strata, villages were sampled with a fixed 1/3 probability, with a floor of one village per strata. After meeting the condition of one village per strata, the floor was taken in the remaining strata, producing a total of

[1]Optimal design standardizes the effect size by dividing the coefficient on the treatment indicator by its standard deviation. Note, optimal design assumes one treatment and one control condition with N/2. To obtain appropriate sample sizes for four conditions, the estimates were doubled.

Figure 2: Power Analysis

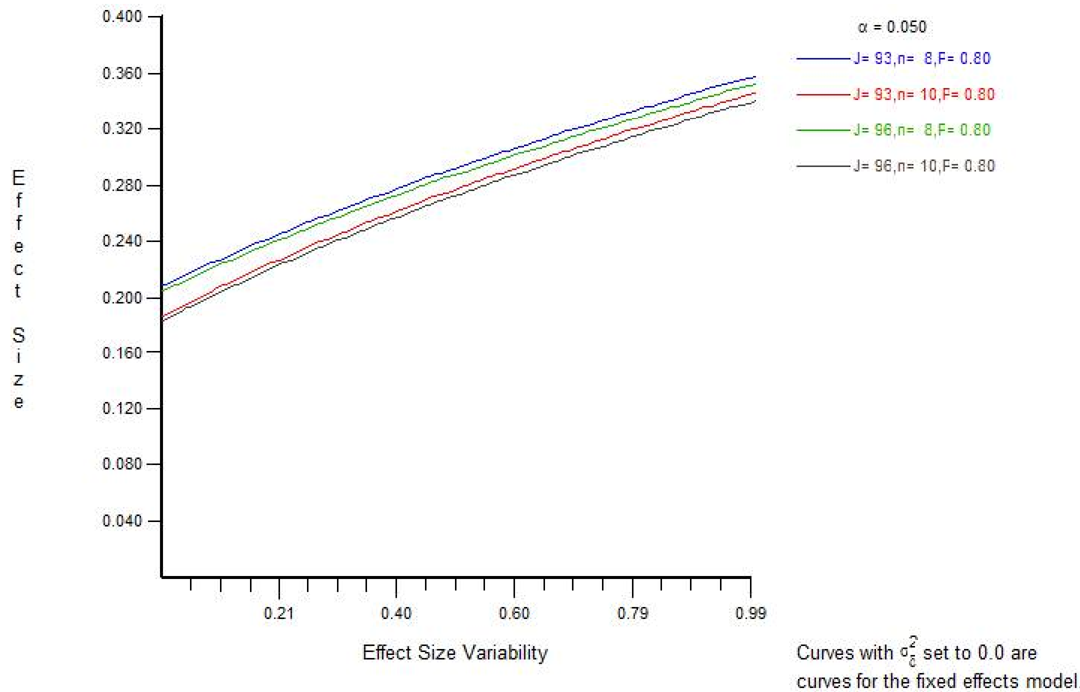
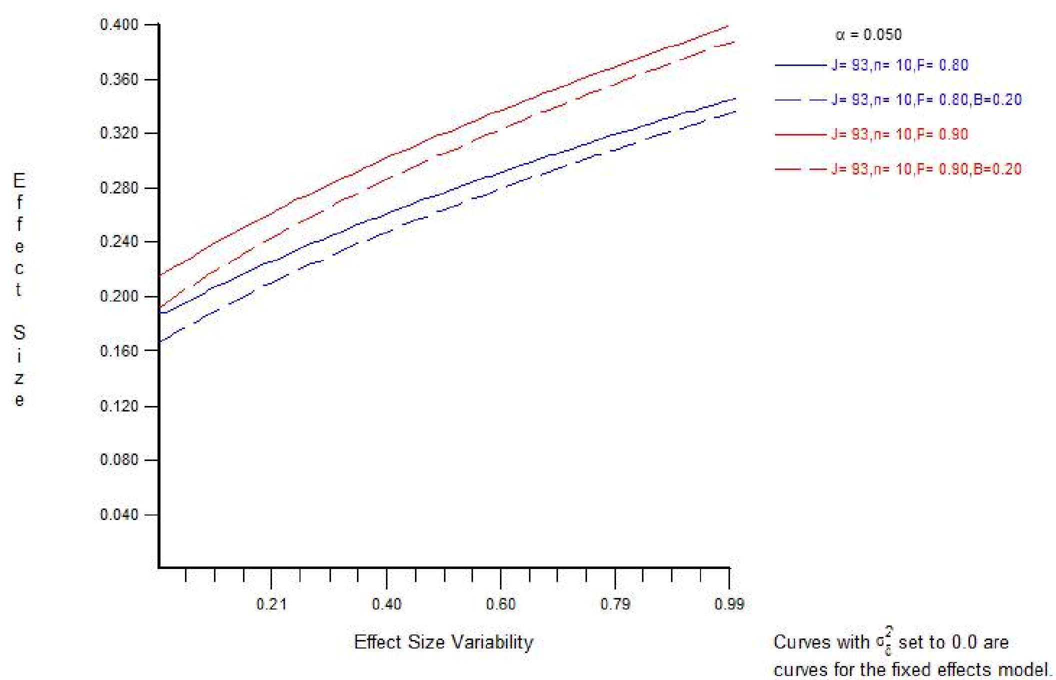


Figure 3: Power Analysis



96 villages. To reach a target of 93, three villages were dropped by identifying the three strata with the highest selection probabilities and reducing the target by one village in each. Table 1 shows basic population information for Blora.

Table 1: Blora Population Information

Subdistrict	**Villages**	**Urban Villages**	**Dusun**	**Households**	**Population**	**Voters (2009)**
Banjarejo	20	0	76	16,268	63,179	40,054
Blora	28	12	58	26,057	93,975	61,298
Bogorejo	14	0	37	7,288	25,053	32,552
Cepu	17	6	38	21,185	75,376	45,082
Japah	18	0	38	10,291	35,617	59,588
Jati	12	0	93	13,868	52,354	22,511
Jepon	25	1	52	17,344	64,431	30,548
Jiken	11	0	39	11,590	37,778	19,590
Kedungtuban	17	0	45	16,446	59,337	47,217
Kradenan	10	0	46	12,741	41,633	72,051
Kunduran	26	1	93	18,906	67,445	46,810
Ngawen	29	2	69	16,588	62,026	35,536
Randublatung	18	2	95	20,912	83,118	27,069
Sambong	10	0	30	7,482	31,046	46,271
Todanan	25	1	77	16,811	64,800	51,079
Tunjungan	15	0	55	13,604	47,469	48,837
TOTALS	**295**	**25**	**941**	**247,381**	**904,637**	**686,635**

4.2 Sampling Dusun

Following the random selection of villages, one dusun/RW was randomly sampled within each village for program implementation. We opted to focus on one dusun in each village for several reasons. Dusun are still large administrative units in Indonesia, each with more than 260 families on average. While spillover is a greater concern at the dusun level than at the larger village level, dusun size was still large enough that piloting did not reveal this to be a major concern. Working at the dusun level also imposed the right 'cost' on respondents who wanted to return their postcards. Leaving one mailbox at the village level was thought to be too costly since villages can be quite big and dispersed, whereas leaving multiple mailboxes at a sub-village level would have required securing permissions from multiple dusun heads and possibly slowed the work of advance and campaign teams. In short, working within one dusun presented the optimal balance between spillover concerns, the costs to respondents associated with the behavioral postcard measure, and implementation demands. Dusun sampling was done by advance teams, which enumerated a list of all dusun in the village with the assistance of the village administration and used a random number table to select one.

4.3 Household Sampling

Within each village/dusun, 20 households were randomly sampled from which one main respondent would be sampled each. In each village, advance teams were able to sample households from the

Buku Induk Penduduk (BIP), which had been updated in December 2009 (two months prior to implementation) to fulfill requirements by the *Dinas Kependudukan dan Catatan Sipil*.[2] Advance teams obtained the complete list of households and used a random number table to select 20. If during implementation a replacement household was needed, the household was selected either by taking the next household on the BIP or, if access to the BIP could not be arranged in time, by visiting the household on the right of the location of the originally sampled household

4.4 Main Respondent Sampling

Main respondents were identified within households by simple random sampling. A household member was eligible if they met the following criteria: 1) They were a member of the household (defined by whether they were on the KK *Kartu Keluarga* or they would be on the KK if it were updated right then [3]; 2) they were 17-65 years of age; 3) they had resided in Blora for at least the previous six months; and 4) they met the gender target for treatment assignment (see Section 4.5). Enumerators made a full list of all members of the household using the first criteria and randomly sampled one member from that list who met criteria 2-4 using a fully randomized table.

Replacement main respondents were the household member listed below the originally sampled main respondent on the sampling table. If a household did not contain any eligible members, then a replacement household was selected following the rules described in Section 4.3.

4.5 Sampling Probabilities

Table 2 shows the levels of sampling and the probabilities attached to each level, where d is the number of dusun in a village, h is the number of households in the village and n is the number of eligible respondents per household.

Table 2: Sampling Weights

	Sampling Level	**Sampling Probability**
A	Village	1/3
B	Dusun	1/d
C	Household	20/h
D	Respondent	1/n

Based on the above, an individuals' probability for selection is based on (A) the probability the village was selected, (B) the probability the dusun was selected, (C) the probability the household was selected, and (D) the probability the respondent was selected. The probability of selection is

[2]We also prepared a compass method, but we did not use it since recently updated household lists were available in all sampled villages.

[3]KKs are typically updated every five years. This accounted for new births or deceased since the last update

represented by:

$$Pr[ABCD] = Pr[A]Pr[B|A]Pr[C|AB]Pr[D|ABC]$$

In other words:

$$1/3 * 1/d * 20/h * 1/n$$

5 Treatment Assignment

Assignment to treatment was done at the individual level and blocked at the village/dusun level. Additionally, treatment assignment was stratified on gender within villages to ensure gender balance. Treatment was assigned to ensure balance across gender and campaigner, both of which could affect the outcome. First, gender targets were set to ensure a balance within each treatment group in each village. Second, a campaign version target was set so that each campaigner implemented each campaign version once per village. Third, treatment was assigned to achieve a gender and version balance for each campaigner every two villages.

Prior to treatment assignment within villages, the order in which villages would be visited was randomly assigned to minimize time effects. The campaign was implemented by three teams. Villages were assigned to the teams non-randomly and according to geographic region (each team visited 31 villages). Villages were grouped into clusters of three proximate villages to make team movement easier. Each team visited a total of 10 clusters (9 clusters of 3 villages and 1 cluster of 4 villages), at the rate of about one cluster per week (they spent two days in each village). Within teams, I randomized the order of the clusters they would visit as well as the order of villages within clusters.

After randomizing village order, assignment to treatment was randomized within villages. With five individuals per cluster arm, it was not possible to achieve a gender balance within each cluster in each village. Targets were set so that, for each team, a gender balance within each treatment group was reached every two villages. For instance, if a target of 2 males was set for version 1 for Village 1, then a target of 3 males would be set for version 1 for Village 2. Next, a target was set to ensure that each campaigner would implement each version once per village (to minimize campaigner effects). Each team had five campaigners and each campaigner consequently completed four visits per village (over two days). Finally, targets were set such that a campaigner achieved balance across treatment groups and genders every two villages. For instance, if Campaigner A implemented version 1 with a male in Village 1, they would do version 1 with a female in Village 2. Once all targets had been set, a random number was generated and observations were ordered 1-20 within each village. These observations were then matched with the list of 20 households that had been randomly sampled by the advance teams. For an example of the treatment assignment document provided to teams for each village, see the sample Village Assignment Sheet.

Table 3 is a regression of the treatment assignment indicator on respondent sex, enumerator, village clusters and the order of villages within clusters. None of these variables significantly predict treatment assignment, showing that treatment assignment was orthogonal.

Table 3: Regression of Treatment Assignment on Predictors

Variable	Coefficient (Std. Err.)
sex	-0.001
	(0.052)
enum	0.000
	(0.003)
cluster_order2	0.000
	(0.009)
order_vils_in_cluster	0.001
	(0.030)
Intercept	2.497**
	(0.113)

5.1 Non-compliance

The treatment assignment was done by the Project Investigator and an assignment sheet was prepared for each team for each village (see the sample Village Advance Sheet). As of this writing, there were only two cases of non-compliance with treatment assignment, and both instances where the canvasser implemented the wrong version out of confusion. In both cases the canvasser randomly sampled another respondent in the dusun and completed an extra visit to maintain the sample size per treatment condition.

6 Other Considerations

6.1 Spillover

Spill-over is always a concern in information experiments. During piloting there was anecdotal evidence of spillover, but typically of basic information in the shared sections of the campaign. There is potentially less cause for concern about spillover for the tax treatment since it is based on a behavioral exercise that would be hard to replicate (although the information itself could be spread). Similarly, the information in the information treatment is rather complicated and would be hard to convey in detail. Additionally, campaigners asked respondents not to discuss the campaign with their neighbors until after the program was over in their village.

Nevertheless, because of the spill-over risks, several steps were taken to measure how widespread it might be. In particular we collected data on:

- Population density of the village and the location of the household within the village on the assumption that information will spread more quickly in denser areas.
- Frequency of respondent communication with individuals in other RT, dusun and villages during a two day period.

- Whether respondents had directly heard information about the campaign, and what information they had heard. This is especially insightful for respondents visited on the second day in each village.
- How many extra postcards (left with the dusun head) were picked up and returned. In each village, five extra postcards were left with the dusun head. The dusun head was instructed not to hand out postcards unless dusun members specifically came to request them. Sampled respondents were also told not to advertise the fact that the dusun head had extra postcards. They were asked to direct other dusun members to speak to the dusun head if community members approached them and wanted to learn about the campaign or take part it in.

6.2 Time Effects

The campaign could have a bigger impact as the election draws closer. I account for time effects by randomizing the order in which subdistricts are enumerated and by ensuring there is a balance across all experiment arms in the lead up to the elections.

6.3 Quality Control

Quality control was primarily performed by team Supervisors, who checked every survey for inconsistencies and missing data. Supervisors also performed spot checks within villages to make sure that enumerators were visiting the correct household and that replacements were selected according to protocol. Surveys were also checked by the Principal Investigator and Project Coordinator when they were returned, where inconsistencies and missing information were sometimes further inquired about. In addition to this daily quality control, separate quality control teams revisited randomly selected households in weeks 4 and 8 of implementation.

7 Data Analysis

I will be analyzing data from participation in a postcard campaign and from pre- and post-treatment survey modules. The postcard serves as the behavioral outcome measure of demand for good government. At the end of the campaign, canvassers encouraged participants to participate in a postcard campaign. Each respondent was given a postcard and told the location of a special mailbox in their village where they could deposit it. The postcard asked individuals to indicate whether they were satisfied with the district government and-if they were not-what governance reforms they wanted. Since returning the postcard required respondents to make some small effort, the postcards measure participants' demand for good governance and their willingness to signal it. To learn the impact of the tax and transparency treatments, we compare the return rates and responses across participants who received different versions of the campaign.

An analysis of the campaign impacts will also be done using pre- and post-treatment survey modules. The survey modules were conducted during the same visit as the campaign and modules were identical across all campaign versions. The pre-treatment survey module was conducted immediately after main respondent sampling and will be used to check for balance, for regression controls, and for heterogeneous treatment effects. The pre-treatment survey includes measures on demographics, income and wealth, tax history, public service usage, media access and consumption, opinions on the district government, and political behavior. A short post-treatment module is conducted after the tax treatment that collects data on economic loss, citizenship, and perceptions of fairness and reciprocity. The survey module following the transparency experiment measures their reactions to the information, their post-treatment attitudes towards government, and their voting priorities.

I will primarily use two-sample ttests as well as ANOVA for multiple-groups to analyze results across treatment groups. I will also present results from linear regression with pre-treatment covariates. The assumption under-pinning analysis is that SUTVA was not violated at the individual level (see Section 6.1) but that any regression should account for correlations across individuals within the same village. Since treatment assignment was blocked at the village level, any linear model should adjust standard errors accordingly, either with robust clustered standard errors, random effects (presuming heterogeneous treatment effects across sites) or fixed effects at level two (presuming homogenous treatment effects across sites) (**?**).[4]

It should be noted that there is still a debate over multiple regression with experimental data; it is both widely used and strongly cautioned against. The argument in favor of pre-treatment covariates is that they reduce variance, which results in more precise estimates of the coefficient on the treatment indicator. Others have argued, however, that multiple regression does not ensure against bias since OLS assumes a linear additive treatment effect. When treatment effects are heterogeneous, treatment assignment and the error term could still be strongly correlated and uncorrectable using standard methods, resulting in both biased coefficients and standard errors (Green, 2008; Sekhon, 2007, 2009).

[4]Green and Vavreck (2008) warn that robust clustered standard errors are still biased downwards (when the number of clusters is small) and that GLS with random effects does slightly better that OLS with robust clustered standard errors.

References

Arceneaux, Kevin. 2005. “Using Cluster Randomized Field Experiments to Study Voting Behavior.” *The ANNALS of the American Academy of Political and Social Science* 601:169–179.

Gerber, Alan and Donald Green. 2000. “The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment.” *American Political Science Review* 94(3).

Green, Donald. 2008. “Regression Adjustments to Experimental Data: Do David Freeman’s Concerns Apply to Political Science.” *unpublished* .

Green, Donald and Lynn Vavreck. 2008. “Analysis of Cluster-Randomized Experiments: A Comparison of Alternative Estimation Approaches.” *Political Analysis* 16:138–152.

Sekhon, Jasjeet. 2007. The Neyman-Rubin Model of Causal Inference and Estimation via Matching Methods. In *The Oxford Handbook of Political Methodology*.

Sekhon, Jasjeet. 2009. “Opiates for the Matches: Matching Methods for Causal Inference.” *Annual Review of Political Science* 12:487–508.