Draft Pre-Analysis Plan for "Rural Institutional Innovation: Can Village Courts in Bangladesh Accelerate Access to Justice and Improve Socio-Economic Outcomes?"

Martin Mattsson and A. Mushfiq Mobarak

July 21, 2019

Abstract

This is a pre-analysis plan describing how we plan to analyze the follow-up data from a Randomized Controlled Trial (RCT) evaluating the effect of the Activating Village Courts Bangladesh (AVCB) program. The AVCB program has as its goal to establish functional Village Courts (VC) in rural Bangladesh. Village Courts are anchored in the lowest tier of government and are designed to resolve smaller disputes in a quick and inexpensive way while still having the enforcement capabilities of the state. The analysis of the follow-up data from this research project aims to answer questions about the effect of the AVCB program on VCs functionality, the effect of VCs on dispute resolution as well as the effects of improved access to justice on dispute frequency, trust and communal harmony as well as economic activities dependent on dispute resolution.

Contents

[1]The pre-analysis plan for the Balance of Randomization table was created and registered with the AEA RCT registry before the collection of the baseline data.

1 Introduction

1.1 Introduction and motivation

The rule of law is generally regarded as a necessary condition for economic development. The judiciary, or the system of courts that interpret and apply the laws, is the main institution ensuring that the rule of law is respected and that justice is accessible to all citizens. However, when designing judiciary systems, there is little high-quality evidence to guide the decision-making of policy makers.

In 2006, the Government of Bangladesh legislated a system of Village Courts (VCs) which were given the power to resolve small disputes. These VCs are anchored in the lowest tier of government, the Union Parishad (UP), with the UP Chair acting as the Chair of the court. In theory, these Village Courts resolve small disputes cheaply and with few administrative complications, but in practice, the implementation of the VC system has been poor and usage of the VCs thus far has been low. To address this problem, the Government of Bangladesh - with technical assistance from UNDP and funding from the EU - has launched a program called Activating Village Courts in Bangladesh (AVCB). The AVCB program makes VCs functional by providing materials, human capital support, and training to the UP officials. This means that currently the functional VCs are almost exclusively within UPs that have received the AVCB program. In its first phase, the AVCB program was implemented in 351 UPs. Now, the government has expanded the program to an additional 1,000 UPs. We have taken advantage of the roll out to evaluate the AVCB program using an RCT by randomizing what UPs have received the program in two of Bangladesh's eight geographical divisions. The evaluation is funded by 3ie (Grant number: DPW1/1100) and the UNDP.

1.2 The AVCB program

The Village Court Act of 2006 requires that a VC be composed of an UP chair (who acts as the Chair of the VC) and four jury members nominated by the plaintiff and the defendant. The plaintiff and the defendant each nominate two jury members, of which two (one from the plaintiff and one from the defendant) must be from the twelve-person UP council. These five-person courts take on both criminal and civil cases and have the authority to adjudicate disputes up to BDT 75,000, or approximately USD 1,000. The VCs do not have the ability to adjudicate more serious criminal cases, such as abduction, rape, or murder. VCs can only impose financial punishments and cannot for example send defendants to prison. In disputes that involve women or children, at least one woman has to be appointed to the court. The AVCB program was initiated because the VC structure had been implemented in almost no UPs by 2009—three years after the passing of the Village Courts Act. The AVCB program trains relevant local government officials, provides physical infrastructure in the form of court room furnishings and stationary, and provides a Village Court assistant who helps with the daily administrative tasks of the VCs. The AVCB program also publicizes the VC system through television commercials, fliers, and even open-air theater, informing local citizens of their right to seek justice from the VCs. In 2017, the AVCB program has been rolled out to an additional 1,080 UPs and will be active for the next three years.

1.3 Main research questions

Our main research questions at this stage of the project are the following:

- Does the AVCB program establish functional VCs?
- Do VCs change the institution in which disputes are resolved?
- Do VCs lower the cost of resolving disputes for citizens?
- Do VCs affect the speed of and the satisfaction with the resolution of each counter party?
- If dispute resolution improves, then do VCs have downstream effects on trust, communal harmony and prevalence of dispute?
  - Are these also downstream effects on economic activities especially dependent on dispute resolution?

1.4 Randomization, data and sampling

1.4.1 Randomization

The AVCB program is implemented at the UP level and our randomization also takes place at the UP level. The randomization process started with the Government and the UNDP selecting 267 UPs that were eligible for the AVCB program in Dhaka and Chittagong divisions, out of these it was agreed that 178 would be randomly selected to receive the program while 89 would be randomly selected to not receive the program. The randomization was stratified on the geographical location of the UPs and took place on June 21, 2016.[2]

1.4.2 Sample of UPs

Of the 267 UPs that are in the RCT, 107 were sampled using simple random sampling stratified on geographical location. For each UP in the sample a "replacement UP" was selected so that if this UP could not be surveyed due to unforeseen circumstances another UP in the same geographical area with the same treatment assignment would be surveyed. These 107 UPs were surveyed during the baseline survey, for the follow-up survey another approximately 70 UPs were added (the remaining 35 control UPs and an additional 35 randomly sampled treatment UPs), these UP will be referred to in the pre-analysis plan as the "additional follow-up UPs".

1.4.3 Data sources

In each UP that we surveyed in the baseline, we collected data in four different ways. Since it is not that common for a household to have a dispute in any given year, we wanted to focus on households that had an ongoing dispute or that had a dispute resolved recently. Therefore, we first conducted a short targeting survey with a large number of households (9,675 in total) to identify the households that had been engaged in disputes. We then conducted a longer household survey with a weighted random sample of the households that we had done the short targeting survey with. Furthermore, we interviewed 3 UP representatives and one UP official (UP secretary) in each UP and conducted a review of the administrative documents relating to the VC in each UP.

In the follow-up survey we will survey the same households as we did in the baseline but we will also survey approximately 70 new UPs. In the 70 additional UPs the same information sources will be used: households, UP representatives and officials and administrative documents. But we will not conducted a targeting survey but instead simply interview a simple random sample of households.

Finally, we will digitize administrative data from District Courts (the lowest level in the formal court system). This data will enable us to measure the number of cases brought to the District Court from each UP in our experiment.

1.4.4 Baseline household sampling procedure

In each UP we surveyed in the baseline, we then selected a ward to conduct our survey in. A surveyor then met with a knowledgeable person within the UP and divided the ward in sub-UP areas, normally villages or *paras* (neighborhoods). We then selected 90 households within the sub-UP area randomly and collected basic household characteristics as well as dispute data from them in what we call the targeting survey. After having conducted the targeting survey we then selected households for the full household survey using weighted random sampling where a higher probability of sampling was given to households that either had an ongoing dispute or had resolved a dispute within the last 2 years. An even higher weight was given to households that had an ongoing dispute within the jurisdiction of the VC since we expect some of these disputes to be resolved in the VCs set up by the AVCB program.

1.5 Theory of change

We have a multi-step theory of change for how the AVCB program might affect the lives of citizens in the UPs. This theory of change represents our hypothesis of the program's potential effects on the outcomes we are interested in analyzing over the course of this project, but it should not be interpreted as an exhaustive list of all the program's potential effects.

The AVCB program creates functional VCs The AVCB program improves both the capability of UPs to run effective VCs and the incentives to do so. This leads to effective VCs that can be used by the people in the UPs.

[2] The random assignment was done by Martin Mattsson using Stata's random number generator (the command "runiform").

New functional VCs are used for dispute resolution Before the VCs existed, UP inhabitants involved in disputes could pursue four possible courses of action: i) bring the case to the District Court, ii) use an informal DRM (e.g. *Shalish* which is the Bengali name for the most common type of informal dispute resolution), iii) resolve the dispute without the help of an institution, or iv) leave the dispute unresolved. Therefore, when the VCs emerge as a fifth option, we expect an increase in the share of disputes resolved in VCs and a decrease in the share of disputes handled in ways i) to iv).

Improvement in access and/or quality of justice Dispute resolution in VCs is quicker and cheaper than in the District Courts. Furthermore, resolutions are better enforced than those delivered by the informal justice system (*Shalish*). Swift and inexpensive judgments increase satisfaction with the dispute resolution process. Also, improved enforcement capabilities reduce the risk of the dispute continuing post-judgment.

Reduction in dispute prevalence and improvement in trust and feelings of safety With laws being enforced more stringently, we expect a reduction in unlawful activities and actions. Note that unlawful activity is not simply defined as traditional crimes (theft, assault, etc.), but also as breaches of business agreements, non-payment of loans, etc and other actions that can cause dispute. With better law enforcement and fewer instances of unlawful behavior, we expect less disputes and crime and that feel safer and trust members of their community more.

Increased economic activity As people feel safer and trust community members more we expect an increase their participation in economic activities that require trust or contract-based engagement with other individuals and counter-parties such as risk sharing, informal credit arrangements and joint investments. Furthermore, contractual and business relationships may expand beyond close family networks to networks where contract enforcement was weaker before the existence of the VC. An improved sense of property rights may also increase investment in previously insecure property, such as disputed land.

Improved economic outcomes Ultimately, as new economic opportunities arise, material well-being will increase and poverty will decrease. Furthermore, if informal risk sharing and credit markets expand the stability of material well being over time increases as households have a better ability to handle shocks.

2 Pre-analysis plan strategy

We will create two pre-analysis plans for this project. The first one, for the Balance of Randomization analysis, has already been written and registered with the AEA RCT registry. It can be found in its entirety together with the outcomes of Balance of Randomization analysis in appendix B. The second pre-analysis plan is this one, specifying how we will analyze the follow-up data before this data is collected.

2.1 Domains of analysis

We have organized the pre-analysis plan according to the domains of outcome variables that we will estimate the AVCB program's effect on. We have aggregated the outcome variables into these domains in order to create one index per domain that we will test a hypothesis on using a statistical test. Aggregating outcomes into a smaller number of indexes reduces the number of hypotheses we will test and therefore reduces the need for multiple hypothesis adjustment of the p-values of these tests, see Section 12.1 for details about how we will adjust for multiple hypothesis testing. The domains we have aggregated our outcome variables into are the following:

- Domain 1: Functionality of VCs (Section 4)
- Domain 2a: Choice of DRM among existing disputes (Section 5.1)
- Domain 2b: Access to justice and quality of dispute resolution among existing disputes (Section 5.2)
- Domain 3: Overall effects on dispute resolution system (Section 6)
- Domain 4: Spillovers on district courts (Section 7)
- Domain 5: Effects on trust, communal harmony, disputes and perceptions of crime and disputes (Section 8)
- Domain 6: Downstream economic effects (Section 9)

2.2 Pre-specifying only the "first level" of analysis

Following Olken [2015] we will pre-specify what one could call the "first level" of analysis, in other words the analysis we had in mind when designing the experiment. In general we will not attempt to design a decision tree and make statements such as "if we find result X we will conduct analysis Y" instead we will simply state how we will conduct the test for result X and once we have seen the result we will decide on what further analysis would be interesting to conduct. This strategy will be explained in the publications following this experiment and we will clearly state what parts of the analysis was pre-specified and what parts arose as a result of the first level results. The one exception to this rule is the sequence in which we will test the AVCB program's effects on our pre-specified domains. This sequence is pre-specified and is described in Section 12.1.2.

Despite our best efforts to write a comprehensive pre-analysis plan there might arise "first-level" interesting questions and/or econometric specifications and tests that were not included in the pre-analysis plan. We may report such finding but with the caveat that they are exploratory findings that needs to be replicated independently in order for the p-values to have their standard interpretation.

2.3 A note on priors

In this pre-analysis plan we will describe what our hypotheses are after each specification. These statements should be interpreted as our priors *if* the AVCB problem works as in our theory of change (see Section 1.5) and that the effects are strong enough for our empirical approach to pick them up. The hypotheses should not be interpreted as our unconditional prior.

3 Econometric specification

3.1 Basic specification

Most of the econometric specifications described below will have the following form:

$$outcome_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{1}$$

Where X_i is a vector of control variables. Our preferred estimate will be the estimate from the model in Equation 1 but we will also report estimates from a regression without control variables for transparency.

3.2 Control variables

We will standardize X_i to contain the variables listed below.

3.2.1 UP official level analyses

- Outcome variable at baseline
- Dummy variable for type of UP official
- Age of UP official
- Years of education of the UP official
- Hours per week spent on dispute resolution at baseline
- VC knowledge score at baseline

3.2.2 Household level analyses

- Outcome variable at baseline
- Household size
- Dummy variable for being an owner of agricultural land

- Total area of land owned
- Years of education of household head
- Distance to UP complex
- Distance to District Court
- Household being involved in dispute in the previous 2 years at baseline

3.2.3 UP level analysis

- Outcome variable at baseline
- UP population
- Fraction of randomly selected households having had a dispute within 1 year at baseline
- Distance to District Court
- Region dummy

3.2.4 Observations without baseline data

If we have no baseline data, which will be the case for all the "additional follow-up" observations, we will set the X_i vector to zero and include a dummy variable for if the observation had no baseline data.

3.2.5 Log transformations and observations equal to zero

If we have specified that we will take the log of a variable, but that this variable turns out to have observations that are zero, we will instead use the inverse hyperbolic sine transformation.

3.3 Recasting variables as standard deviations

Before performing our analysis we will recast several of our variables as standard deviations. The process for recasting a variable x as standard deviations of that variable, SD_x, is the following:

$$SD\_x_i \equiv \frac{x_i - \bar{x}}{\left(\frac{1}{N^{control}} \sum_{i=1}^{N^{control}} (x_i - \bar{x}^{control})^2\right)^{\frac{1}{2}}} \tag{2}$$

Where $\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$ such that the nominator is simply the demeaned variable. The nominator is the estimate of the control group standard deviation.

4 Domain 1: VCs functionality

As described above, the first step in the theory of change is that the AVCB program improves the functioning of the VC. We will measure this using five indicators of how well the VC is functioning, three which are collected at the UP level and two which are collected from the household interviews.

4.1 UP-level indicators

UP officials variables:

- Knowledge about VC rules and regulation by UP officials will be measured by the standard deviation of the number of correct answers to a quiz. The topics of the quiz stay the same but the exact questions of the quiz changes between the baseline and the follow-up survey:

$$SD\_Score_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{3}$$

- Number of hours per week UP officials state spending on resolving cases in VCs:

$$ln\_VC\_hours_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{4}$$

UP administrative records variables:

- Fraction of VC documentation protocols followed by the UP:

$$Percent\_protocols_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{5}$$

4.2 Household level indicators

Household variables

The first and most basic measure of how functional the VC is from a household perspective is if the households themselves claim that there is an active VC in the UP. This will be measured using the following regression:

$$Answer\_yes\_VC\_active_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{6}$$

The other household variable we will use is the households stated inclination of using the VC to resolve four hypothetical disputes within the jurisdiction of the VC. The four disputes are:

- BDT 10,000 loan not repaid
- Physical assault on family member so bad they needed medical treatment
- Illegal occupation of land
- Intentional damage to crop

The number of cases that each respondent states she would resolve using a VC is recorded and the following regression will estimate how the AVCB program has affected this variables:

$$Number\_resolved\_VC_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{7}$$

For all of the regressions in Section 4 our prior is that $\beta > 0$.

5 Domain 2: Effect on existing disputes (dispute level outcomes)

5.1 Domain 2a: Choice of Dispute Resolution Mechanism (DRM)

As per the theory of change above, the next question we will be asking is how the AVCB program changes the way that households resolves disputes.

Using a multinomial logistic regression at the preexisting dispute level, (i.e. i represents a specific dispute at baseline) we will investigate how the AVCB program affects what institutions people actually use when resolving a dispute. The following multinomial logit model will be estimated using maximum likelihood:

$$p\left(resolution_i = j | Treat_i, X_i\right) = \frac{exp\left(\alpha_j + \beta_j Treat_i + \gamma_j X_i + \varepsilon_i\right)}{1 + \sum_{h=1}^{H} exp\left(\alpha_j + \beta_j Treat_i + \gamma_j X_i + \varepsilon_i\right)} \tag{8}$$

Where the $Resolution_i$ variable can take the following values:

- $Resolution_i = 0$ if the dispute is unresolved[3]
- $Resolution_i = 1$ if the dispute was resolved in a VC
- $Resolution_i = 2$ if the dispute was resolved in a district court
- $Resolution_i = 3$ if the dispute was resolved in a Shalish
- $Resolution_i = 4$ if the dispute was resolved without any of these three institutions

X_i is a vector of baseline characteristics such as type of dispute, value of dispute, etc. (Here we have to think about a systematic way of defining X_i that maximizes power.)

Our prior is that $\beta_{VC} > 0$ while the coefficients for the other variables are zero or negative.

[3] $p\left(resolution_i = 0 | Treat_i, X_i\right) = \frac{1}{1 + \sum_{h=1}^{H} exp(\alpha_h + \beta_h Treat_h + \gamma_h X_i + \varepsilon_i)}$

5.2 Domain 2b: Access to justice and quality of dispute resolution

5.2.1 Access to justice

Our next question will be if the AVCB program has increased access to justice measured as the cost of resolving an preexisting dispute (i.e. the typical dispute before the AVCB program was implemented). This is admittedly a very narrow definition of "access to justice", since the access to justice concept could also include knowledge about DRM and so on. However, the cost measure be measured quantitatively in a consistent manner and is probably one of the main barriers for the population to access DRM in our context.

We will start this analysis by calculating a cost measure of each dispute resolution in terms of fees, cost of legal counsel, the cost of travel to and from the hearings, the opportunity cost of time spent on resolving the dispute as well as any additional payments made to anyone with regards to the dispute resolution, such as speed-money (a form of bribe) paid to court clerks.

$TotalCost =$ fees + cost of counsel + cost of travel + days spent on resolution $\times$ average daily wage + other costs

The cost variable will be winsorized at the 99th percentile and we will then estimate the effect of AVCB program on the average cost per dispute:

$$ln\_cost_i = \alpha + \beta Treat_i + \varepsilon_i \tag{9}$$

Our prior is that $\beta < 0$.

5.2.2 Quality of dispute resolution

We will also measure the quality of dispute resolution in three ways. The first is the way the respondent describes the relationship with the other party in the dispute (5 point scale) ($relationship_i$), the second is the respondents subjective satisfaction with the court process ($satisfaction_i$) and the third is the time the dispute resolution took from first approaching the dispute resolution institution to the enforcement of the decision (for the cases that are not yet resolved we will use expected date of enforcement) ($process\_days_i$). We will analyze each of these variables but our preferred specification is to convert the answers to these three questions into standard deviations and adding them together creating a "Quality of resolution index"[4] and then run the following regression:

$$\text{Quality of resolution index}_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{10}$$

Our prior is that $\beta > 0$.

5.2.3 Stress among those with disputes

We will measure stress levels using the Perceived Stress Scale (PSS) [Cohen et al., 1983] (translated into Bengali) and run the following regression:

$$SD\_PSS_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{11}$$

Our prior is that $\beta < 0$, i.e. that individuals having access to a VC to resolve their dispute experience less stress.

6 Domain 3: Overall effects on dispute resolution

6.1 Number of disputes, number of unresolved vs. resolved disputes

After having studied the effect of the AVCB program on preexisting disputes it is natural to ask the question, who does the AVCB program affect overall dispute resolution taking into account the possibility that disputes themselves might be affected by the program. Our starting point will be to measure how the AVCB program affects the number of disputes that exists or have existed since the baseline survey using the following 3 regressions:

$$\#\_disputes = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{12}$$

[4] Quality of resolution index$_i = SD_relationship_i + SD_satisfaction_i + SD_process_days$

$$\#\_unresolved\_disputes = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{13}$$

$$\#\_resolved\_disputes = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{14}$$

Regression equation 12 above measures the effect of the AVCB program on the log total number of disputes experienced since the baseline survey, while the next 5 regressions measures the outcome (unresolved, resolved in VC, resolved in VC, etc.) of these disputes. It is possible that the AVCB causes more disputes to be resolved but also causes more disputes to occur and therefore lead to an increase in both resolved and unresolved disputes. We do not have any strong priors for the signs of the betas of these regressions and we will therefore not include them in the Domain 3 index.

6.2 Number of disputes resolved by DRM

We will then study where the disputes were resolved using the following 4 regressions:

$$\#\_relsolved\_VC = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{15}$$

$$\#\_relsolved\_DC = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{16}$$

$$\#\_relsolved\_Shalish = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{17}$$

$$\#\_relsolved\_other = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{18}$$

Our hypothesis is that $\beta > 0$ for $\#\_relsolved\_VC$ but for the other betas we have no strong priors we will therefore only include $\#\_relsolved\_VC$ in the Domain 3 index of outcomes.

6.3 Effect on overall assessment of justice system

We will also test if the AVCB program improves households overall Satisfaction with the justice system they feel that they have access to using the following regression:

$$SD\_satisfaction_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{19}$$

Our prior is that $\beta > 0$.

7 Domain 4: Spillovers onto the number of disputes going to District Courts

One of the stated goals by the Government and the UNDP is for the VC system to reduce the burden on the existing formal court system. To assess if functional VCs leads to fewer cases being brought to the district courts we will collect administrative data from district courts on the number of cases that have been received from treatment and control UPs. We will then run the following regression at the UP level:

$$ln(\#\_of\_cases_u) = \alpha + \beta Treat_u + \gamma X_u + \varepsilon_u \tag{20}$$

If one of the benefits from the AVCB program is that it decreases the workload on the District Courts we expect $\beta < 0$.

8 Domain 5: Effects on trust, communal harmony, disputes and perceptions of crime and disputes

8.1 Trust: survey measures

We will test if the AVCB program affects the commonly used binary answer question "Can most people can be trusted or you can't be too careful in dealing with people?". We will estimate the effect on the answer to this question using the following logistic regression:

$$p\left(trust_i^{dummy} = 1 | Treat_i, X_i\right) = \Lambda\left(\alpha + \beta Treat_i + \gamma X_i + \varepsilon_i\right) \tag{21}$$

We will also estimate the effect on the measure "How much do you trust each of the following types of people: relatives, neighbors, other people you know" (on a scale from 1 to 5) by adding up the standard deviations of the three answers and running the following regression:

$$sum\_SD\_trust_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i$$

Our prior is that $\beta > 0$ in both regressions.

8.2 Communal harmony

We will test if the AVCB program effects a communal harmony be estimating the effect on the answers to the question: "How much harmony or conflict exists between you and your 5 closest neighbors?" with the following regression:

$$SD\_harmony_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i$$

Our prior is that $\beta > 0$.

8.3 Perceptions of crime and unresolved disputes

We will test if the AVCB program affects the perceptions of how big of a problem crime and disputes are in the village by estimating the effect on the answers to the questions: "How big of a problem crime is in your village?" and "How big of a problem unresolved disputes are in your village?" with the following regressions:

$$SD\_perception\_crime_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i$$

$$SD\_perception\_unresolved\_disputes_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i$$

Our prior is that$\beta < 0$.

8.4 Frequency of dispute

We will create a dummy variable $dispute^{dummy}$ for if you have been the victim of a crime or had a dispute started in the previous year run the following binary logistic regression at the individual level:

$$p\left(dispute_i^{dummy} = 1 | Treat_i, X_i\right) = \Lambda\left(\alpha + \beta Treat_i + \gamma X_i + \varepsilon_i\right) \tag{22}$$

If the VC are effective in deterring crime and the start of disputes $\beta < 0$.

9 Domain 6: Downstream economic effects

Our strategy for estimating the effect on economic behavior will be threefold. The first is to simply ask households how many economic agreements they entered into in the year, the next two strategies are to measure the effect in two economics activities that we think may be directly affect by the fear of a breach of contract namely joint investments and informal credit.

9.1 Number of disputes and total resources spent on dispute resolution

We will also measure the effect the AVCB program has on the number of disputes taken to any type of dispute resolution institution and the total number of resources that is spent on dispute resolution. For that we will run the two following household level regressions.

$$\text{Number of disputes brought to a dispute resolution institution}_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{23}$$

$$\text{Monetary value of resources spent on dispute resolution}_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{24}$$

Note that theoretically we do not have a prior for AVCB program's effect on the two variables above. Even if the AVCB program is successful there may be an increase in the number of disputes brought to a dispute resolution since dispute resolution is now more accessible and have higher quality, the same goes for the resources spent on dispute resolution.

9.2 Amount and counterparts of economic agreements

Our first strategy will be to ask households how many economic agreements, how large and with whom they entered into in the previous year. We will also analyses the total value of all contracts (winsorized at the 99th percentile) that occurred during that previous year using the following regression.

$$\text{Total value of contracts} = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{25}$$

We will then classify the people they entered into agreements into as family and non-family.

We will estimate what fraction of a households economic agreement that are outside the immediate family using the following regression:

$$\text{Fraction of agreements value that are outside family} = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{26}$$

We will estimate the faction of agreements that are new, i.e. that the household has not done any economic agreements with this counterpart before:

$$\text{Fraction of agreements value that are new} = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{27}$$

We will estimate the faction of agreements that are with people from outside the village:

$$\text{Fraction of agreements value with counterparts outside the village} = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{28}$$

9.3 Joint investments

Joint investments is one of the economic activities that we expect may be affected by better contract enforcement. To investigate this hypothesis we will collect data on all the investments that households have made in the previous year and how many of these we made jointly with someone else. We will investigate both the effect on total investments (winsorized at the 99th percentile) and the fraction of joint investments.

$$\text{Total amount of investments} = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{29}$$

$$\text{Fraction of investments that were joint} = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{30}$$

9.4 Risk sharing

If improved communal harmony increases the value of being part of the community, we expect to see more risk sharing and therefore more gifts and informal credit. We will measure risk sharing in two ways.

9.4.1 Gifts

Our first measure will be to ask about the monetary value of gifts the household has actually given and received since the last year:

$$\text{Total amount of gifts given} = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{31}$$

$$\text{Total amount of gifts recieved} = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{32}$$

Our hypothesis is that both $\beta > 0$.

9.4.2 Credit markets

Another important source of risk sharing is the informal credit market. We expect lending and borrowing to be affected by better contract enforcement. If enforcement of credit contracts between friends, family and neighbors is improved we expect to see more lending, lower interest rates and fewer loans being reneged on.

To investigate these hypotheses we will collect data on all the outstanding lending and borrowing of each households as well as the new lending and borrowing a household has done over the previous year. We will begin by running the following regressions for the amount of total credit that is outstanding in the credit market:

$$\text{Total amount of outstanding lending} = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{33}$$

$$\text{Total amount of outstanding borrowing} = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{34}$$

Our hypothesis is that both $\beta > 0$.

10 Other outcomes

10.1 Satisfaction with UP chair

We estimate whether and how the AVCB program effects the satisfaction with the UP chair by estimating the effect on the answer to the question "Do you think the UP chair has done a good job looking after the needs of your village?"

$$SD\_UP\_chair\_satisfaction_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{35}$$

Since we do not have a clear hypothesis in which direction the AVCBs will affect satisfaction with the UP chair we will not include it in any of our aggregated measures.

10.2 Stress

To assess whether the VCs affect the average level of stress in the population, respondents will take the standard Perceived Stress Scale (PSS) [Cohen et al., 1983] (translated into Bengali) and run the following regression:

$$SD\_PSS_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{36}$$

Since we do not have a clear hypothesis in which direction the VCs will affect the PSS we will not include it in any of our aggregated measure aggregated measures.

11 Heterogeneous effects

11.1 People with disputes at baseline

We expect the benefits of the AVCB program to a household to be positively related to the household's propensity to have disputes. We expect that the best predictor of having had a dispute in the period between the baseline and the follow-up survey to be if the household either had had a dispute within 2 years before the baseline or if it had an ongoing dispute at the baseline. Hence we will conduct heterogeneous effect analyses for all household level outcomes using the following regression:

$$Outcome_i = \alpha + \beta_1 Treat_i + \beta_2 Treat_i \times BaselineDispute_i + \gamma X_i + \varepsilon_i \tag{37}$$

Our prior is that β_2 is positive when our prior is that the overall effect is positive and that it is negative when our prior on the overall effect is negative.

11.2 Political bias and nepotism in the VCs

A common criticism against the VC system is that the UP chairs are a politically elected persons that may be more interested in providing patronage to their political supporters than to provide fair judgments in disputes. We will test if there is any empirical evidence supporting this criticism by estimating the heterogeneous effect of the AVCB program on households that are supports the same political party as the UP chair. We will conduct these heterogeneous effects test for both outcomes at the household and existing dispute level.

$$Outcome_i = \alpha + \beta_1 Treat_i + \beta_2 Treat_i \times SameParty_i + \gamma X_i + \varepsilon_i \tag{38}$$

Another problem of providing the UP chairs extended powers to provide judgments in disputes is that they might use these powers to provide favors for people who are close to them but who do not necessarily support the same political party. We will test this by creating an index for how well people know the UP chair using the following questions:

- How well do you know the UP chair? (Scale 1 to 5)
- If you brought up one of your problems with the UP chair, how hard would he/she work to try to solve it? (Scale 1 to 5)
- Do you think the UP chair has done a good job looking after your personal needs? (Scale 1 to 5)

Using the standard deviations of these answers we are will to create an index for closeness with the UP chair that can be interacted with the treatment variable in the following way:

$$Outcome_i = \alpha + \beta_1 Treat_i + \beta_2 Treat_i \times index_i + \gamma X_i + \varepsilon_i \tag{39}$$

We will use this specification to investigate heterogeneous effects for all the household level and existing dispute level outcomes described above.

12 Adjustments for multiple hypothesis testing and unbalanced attrition

12.1 Multiple hypothesis testing

As is evident from the many analyses described above, there is a high risk of false positives is all estimates significantly different from zero are automatically treated as "the effect of the intervention". To minimize the risk for spurious results due to "chance" and for our p-values to have the standard interpretation we will define domains of related outcome variables for which we will create summary indexes following Anderson [2008]. Within each domain we also analyze each component of the domain while adjusting the p-values for multiple hypothesis-testing. Furthermore, we will analyze the domains in a pre-determined sequential way to avoid having to adjust the p-values of the standardized treatment effects for multiple hypothesis testing.

12.1.1 Creating summary indices

Within each domain with multiple outcome variables (all except Section 7 and Section 5.1) we will calculate a standardized treatment effect by creating an index. We will do this by following the 4 steps below:[5]

1. We will change signs of the outcome variables so that a higher number is always the direction of our prior. Variables for which we do not have a prior we will not enter into the indices.
2. Convert all outcome into standard deviations as described in Section 3.3. I will denote the vector of standardized outcomes in domain j $\tilde{y}_j$.
3. Create a vector of weights for each outcome within a domain equal to the sum of the row entries of the inverted covariance matrix for all variables within that domain, normalized by dividing it by the sum of all weights. Formally, $w_j = \left[\left(1'\hat{\Sigma}_j^{-1}\right)\left(1'\hat{\Sigma}_j^{-1}1\right)^{-1}\right]'$ where $\hat{\Sigma}_j$ is the covariance matrix of all variables in domain j and $\mathbf{1}$ is a column vector of ones.
4. Create an index that is equal to the weighted average of the standard deviations of the outcome variables. Formally $Index_{ij} = w_j^{'}\tilde{y}_{ij}$

After having generated these indices we will run the following regression to test the hypothesis that the treatment has no effect on any of the variables:

$$Index_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{40}$$

This procedure has the advantage of reducing the number of hypothesis tests within a domain into one test and thereby avoiding the need for multiple hypothesis testing adjustments within a domain. It also increases our statistical power if our theory of change is correct, since several insignificant results can add up to one significant result if many of the insignificant estimates are (or more specifically, their weighted average) in the same directions as our prior.

12.1.2 Sequence of analysis

As outlined in our theory of change, if there is no effect of the AVCB program on the functionality of VCs (Domain 1), we do not expect there to be any effect on more downstream outcomes such as the other domains. Hence we will analyze the effect on the the functionality of VCs first and if there is no effect there we will interpret effects on other domains very cautiously and as exploratory results that needs to be replicated for the p-values to have their standard interpretation.

Similarly, we do not expect the access to justice and quality of dispute resolution to change unless the actual choice of DRM changes. Hence, if there is no effect on the actual choice of DRM (Domain 2a) we will treat any effect on the access to justice and quality of dispute resolution (Domain 2b) very cautiously and as exploratory results.

Domain 3 (Overall effects on justice system), Domain 4 (Spillovers on district courts) and Domain 5 (Effects on trust, communal harmony, disputes and perceptions of crime and disputes) will be analyzed separately from Domains 2 and 3 as these could be affected even if preexisting disputes were not affected by the AVCB program. Since we are testing the effect on three domains in parallel we will report the family-wise error rate adjusted p-values (as per the procedure in Section 12.1.3) for these 3 outcomes along with the standard p-values.

Following Finkelstein et al. [2012] we will not adjust p-values for multiple hypothesis testing across conceptually distinct analyses, in our case the effect of the AVCB program on preexisting disputes (Domain 2a and 2b) and the effect of the AVCB program on our sample (Domain 3, 4, 5 and 6).

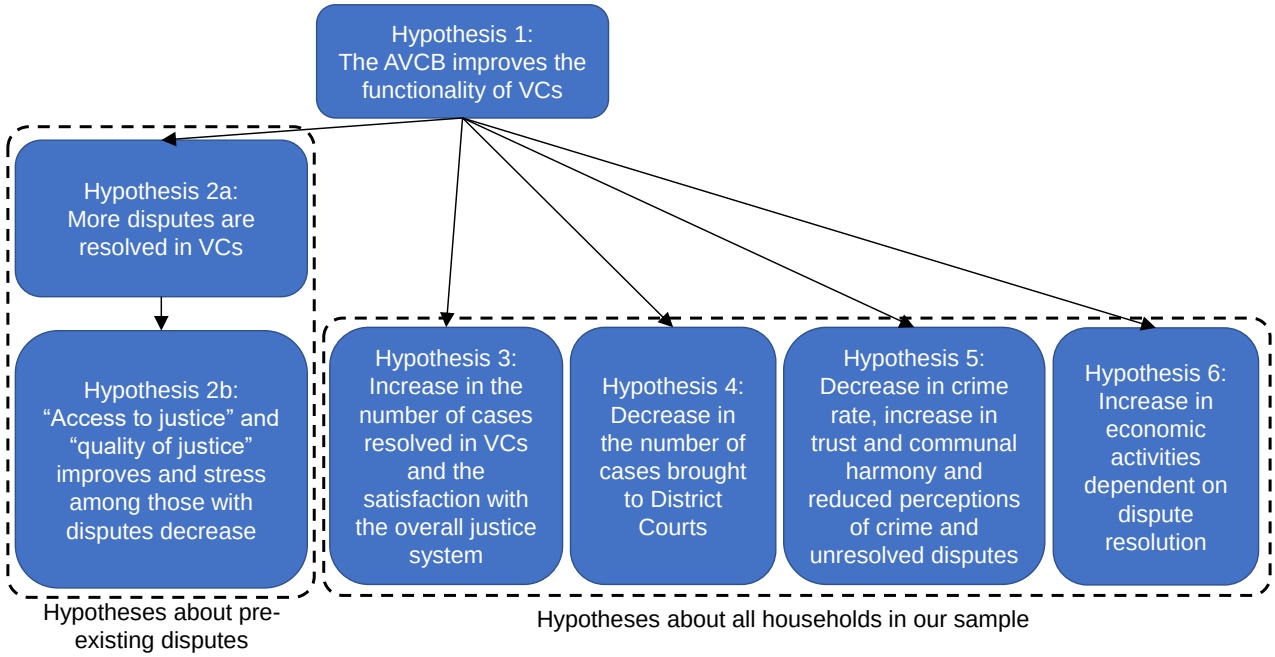
12.1.3 Family-wise error rate adjusted p-values within domain

Within each domain, and across the indices for domains 4, 5 and 6, we will calculate the Family-Wise Error Rate (FWER) adjusted p-values based on 10,000 iterations of Anderson [2008] version of the free step-down re-sampling method of Westfall and Young [1993]. The algorithm has the following 7 steps for a family of M outcomes:

1. Sort outcomes in order of increasing p-values of the rejection of the null hypothesis, $y_1, y_2, ..., y_M$ such that $p_1 < p_2 < ... < p_M$

[5] Adopted from Anderson [2008].

Figure 1: Sequence of analysis



2. Simulate a data set where the null hypothesis is true by drawing simulated treatment assignments from the actual distribution of the treatment assignment without replacement.

3. Calculate a set of simulated p-values $(p_1^*, p_2^*, ..., p_M^*)$ for the rejection of the null for each outcome using the simulated treatment assignments. These p-values will not have the same monotonicity as $p_1 < p_2 < ... < p_M$.

4. Enforce the original monotonicity by setting $p_r^{**} = min\left\{p_r^*, p_{r+1}^*, ..., p_M^*\right\}$ for all r, where r represents the number in the original ordering. Now, by construction, $p_1^{**} \leq p_2^{**} \leq ... \leq p_M^{**}$.

5. Perform 10,000 replications of steps 2-4 and count S_r, the number of times that $p_r^{**} < p_r$, for each r.

6. Generate $p_r^{FWER*} = \frac{S_r}{10{,}000}$

7. The Family-Wise Error Rate adjusted p-values will be $p_r^{FWER} = min\left\{p_r^{FWER*}, p_{r+1}^{FWER*}, ..., p_M^{FWER*}\right\}$

12.2 Attrition

We will report attrition between baseline and follow-up surveys for both the treatment and control group. We will also test for if attrition was affected by the treatment using the following regression:

$$Attrition_i = \alpha + \beta Treat_i + \gamma X_i + \varepsilon_i \tag{41}$$

Furthermore, since attrition can be of a different nature in the treatment and control group even if it is on average the same, we will restrict the sample to the attriting households and run the analysis as we did for the baseline balance check (see Section B). If either the β in regression equation 41 is significantly different from zero or if the balance check of the attriting households rejects the statistical tests described in Section B.2 we will use a Heckman selection correction Heckman [1979] as well as Lee bounds Lee [2009] to assess the extent to which the attrition may effect our estimates of the effect of the AVCB program.

References

Michael L Anderson. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103 (484):1481–1495, 2008.

Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. A global measure of perceived stress. *Journal of health and social behavior*, pages 385–396, 1983.

Amy Finkelstein, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group. The oregon health insurance experiment: evidence from the first year. *The Quarterly journal of economics*, 127(3):1057–1106, 2012.

J Heckman. Sample selection bias as a specification error. *Econometrica, 47 (1)*, 1979.

David S Lee. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102, 2009.

Benjamin A. Olken. Promises and perils of pre-analysis plans. *Journal of Economic Perspectives*, 29(3):61–80, 2015.

Peter Westfall and Stanley Young. *Resampling-Based Multiple Testing*. New York: Wiley, 1993.

A Appendix: List of of acronyms

- Activating Village Court Bangladesh (AVCB)
- Dispute Resolution Mechanism (DRM)
- Family-Wise Error Rate (FWER)
- Perceived Stress Scale (PSS)
- Randomized Controlled Trial (RCT)
- Union Parishad (UP)
- Village Court (VC)

B Pre-analysis plan for Balance of Randomization analysis[6]

In our balance of randomization table we will show that the treatment and control UPs are balanced on the means of the main outcome variables as well and standard demographic variables. We have selected variables into the balance of randomization table keeping two rationales in mind. The first is that the main outcome variables should be included since if these are unbalanced that may raise questions about the main results. The second is to include variables that are important to give the reader a overview of our study population such as age, household size and income.

B.1 Variables for the balance of randomization tests

Household level variables:

- Average age in household (in years)
- Number of individuals in household
- Natural logarithm of per capita household income (winsorized at the 99th percentile)
- Yes to the question if there is a functional VC in the UP (binary variable)

[6]The pre-analysis plan for the Balance of Randomization table was created and registered with the AEA RCT registry before the collection of the baseline data.

- Putting VC as the institution where to resolve hypothetical disputes within the VC jurisdiction (sum of 3 binary variables)
- Total number of disputes (ongoing or resolved within the two years)[7]
- Number of disputes resolved using Shalish (within the last two years)
- Number of disputes resolved using VC (within the last two years)
- Number of disputes resolved using district court (within the last two years)
- Number of disputes resolved without an institution (within the last two years)
- Number of unresolved disputes
- Total amount of resources spent on dispute resolution within the last year (winsorized at the 99th percentile)
- Percentage of households that have been the victim of a crime (within the last two years)
- Total amount of investment within the last year (winsorized at the 99th percentile)
  - of which: was made jointly with a business partner
- Total outstanding amount of money lent out to friends and family (winsorized at the 99th percentile)[8]
- Total outstanding amount of money borrowed from friends and family (winsorized at the 99th percentile)[9]
- Satisfaction with overall justice system (Scale 1 to 5)
- Satisfaction with the UP chair (Scale 1 to 5)

UP level variables:

- UP population (number of households)
- Fraction of randomly selected households with a dispute (ongoing or resolved within last year)
- What fraction of the VC records did the UP keep updated
- What fraction of the knowledge question on how to conduct a VC could the UP chair respond to correctly?
- What fraction of the knowledge question on how to conduct a VC could the UP members respond to correctly?
- Number of cases from UP that has gone to the district court (Number collected from district court's administrative records)

B.2 Statistical tests and interpretation of balance of randomization results

In the balance of randomization table we will test for equality of means between the treatment and control groups by running the following regression for each outcome variable:

$$outcome_i = \alpha + \beta treat_i + \varepsilon_i$$

Where the null hypothesis will be that $\beta = 0$. The standard errors will be heteroskedasticity robust and clustered at the UP level.

In addition we will perform two tests for joint orthogonality, one for the household level data and one for the UP level data, by running the following linear OLS regressions:

$$treat_i = \alpha + \beta_1 outcome.1_i + \beta_2 outcome.2_i + ... + \beta_k outcome.k_i + \varepsilon_i \tag{42}$$

[7] In the original version of the pre-analysis plan the time frame was described as one year. However, due to changes in the final version of the questionnaire, the time frame in the balance of randomization was changed to two years.

[8] The winsorization of this variable had been forgotten in the original pre-analysis plan and was added after the baseline data had been collected

[9] The winsorization of this variable had been forgotten in the original pre-analysis plan and was added after the baseline data had been collected

We will then use an F-test to test the joint null hypothesis $\beta_1 = \beta_2 = ... = \beta_k = 0$. As long as less than 4 individual tests rejects the null at the 95% significance level (the probability of 4 or more tests rejecting the null when the null is true is 0.026) and that neither of the tests for joint orthogonality reject the null hypothesis at the 95% level, we will interpret the results as the treatment and control areas being balanced.

If our result is that the treatment and control areas are balanced we will proceed with the analysis as outlined in the pre-analysis plan that we will submit after having analyzed the baseline data. If we it turns out that by change the treatment and control areas are not balanced we will control for all the baseline values of the variables that are unbalanced in the baseline data through out our analysis.

B.3 Potential changes of the variables in the balance of randomization table

Since the pre-analysis plan in this section was written before the baseline data was collected there may be some changes to the variables in the balance of randomization table depending on what outcome variables we choose to use in our full pre-analysis plan and final paper. In the interest of full transparency, if any variables are dropped from the original list, we will publish the results of the statistical tests described above using the original list of variables as a footnote to the balance of randomization table.

B.4 Results of Balance of Randomization analysis

Table 1 and Table 2 show the result of the balance of randomization test for Household and UP level variables respectively. Among the 23 variables only one (number of disputes resolved in other ways than *shalish*, DC and VC) is significantly different between the treatment and control UPs at the 95% significance level. Furthermore, neither of the two F-test for joint significance could reject the null hypothesis that none of the variables can predict the treatment assignment at the 95% significance level.

Table 1: Balance of Randomization: Household level variables

Variable	(1) Control UPs N/[Clusters]	Mean/SE	(2) Treatment UPs N/[Clusters]	Mean/SE	T-test Difference (1)-(2)
Average household age	1618	28.774	1588	28.375	0.399
	[54]	(0.372)	[53]	(0.385)	
Number of household members	1618	5.501	1588	5.458	0.043
	[54]	(0.087)	[53]	(0.092)	
Per capita income (log)	1570	10.220	1542	10.257	-0.036
	[54]	(0.036)	[53]	(0.043)	
Is there a functional VC in UP?	1618	0.084	1588	0.069	0.015
	[54]	(0.014)	[53]	(0.011)	
Out of 4 hypothetical disputes,					
number resolved in VC	1618	0.046	1588	0.051	-0.005
	[54]	(0.011)	[53]	(0.013)	
Total disputes (past 2 years)	1618	0.515	1588	0.491	0.024
	[54]	(0.036)	[53]	(0.037)	
Disputes resolved in Shalish (past 2 years)	1618	0.114	1588	0.113	0.000
	[54]	(0.013)	[53]	(0.013)	
Disputes resolved in DC (past 2 years)	1618	0.027	1588	0.035	-0.009
	[54]	(0.005)	[53]	(0.007)	
Disputes resolved in VC (past 2 years)	1618	0.003	1588	0.003	0.001
	[54]	(0.002)	[53]	(0.001)	
Disputes resolved in other ways (past 2 years)	1618	0.020	1588	0.038	-0.018**
	[54]	(0.004)	[53]	(0.007)	
Number of unresolved disputes	1618	0.352	1588	0.302	0.049
	[54]	(0.030)	[53]	(0.030)	
Total cost of dispute resolution (past 2 years)	1618	3073.525	1588	2306.745	766.780
	[54]	(402.031)	[53]	(308.167)	
Has been the victim of a crime (past 2 years)	1618	0.255	1588	0.220	0.034
	[54]	(0.018)	[53]	(0.015)	
Total amount of investment (last year)	1618	39519.685	1588	38685.673	834.012
	[54]	(3077.423)	[53]	(4068.588)	
Amount invested jointly with someone else	1618	262.052	1588	143.577	118.475*
	[54]	(49.403)	[53]	(37.712)	
Total outstanding amount lent	1618	3958.158	1588	4180.730	-222.572
	[54]	(469.983)	[53]	(602.109)	
Total current debt	1618	46898.896	1588	46659.384	239.512
	[54]	(3561.751)	[53]	(2933.483)	
Satisfaction with justice system (1 to 5)	1599	2.279	1562	2.245	0.034
	[54]	(0.026)	[53]	(0.038)	
Satisfaction with VC chair (1 to 5)	1618	3.631	1588	3.697	-0.066
	[54]	(0.071)	[53]	(0.056)	
F-test of joint significance (F-stat)					1.692*
F-test, number of observations					3067

Notes: The value displayed for t-tests are the differences in the means across the groups. The value displayed for F-tests are the F-statistics. Standard errors are clustered at variable union. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 2: Balance of Randomization: UP level variables

	(1) Control UPs		(2) Treatment UPs		T-test Difference
Variable	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	(1)-(2)
UP population (number of households)	54	6305.833	53	5878.283	427.550
	[54]	(397.843)	[53]	(454.967)	
Fraction of households with disputes	54	0.158	53	0.158	-0.000
	[54]	(0.013)	[53]	(0.012)	
Fraction of VC records kept	54	0.080	53	0.067	0.013
	[54]	(0.013)	[53]	(0.019)	
VC knowledge test score (UP chair)?	54	0.131	53	0.123	0.008
	[54]	(0.007)	[53]	(0.007)	
VC knowledge test score (Other UP officials)?	54	0.294	53	0.294	0.001
	[54]	(0.012)	[53]	(0.015)	
F-test of joint significance (F-stat)					0.277
F-test, number of observations					107

Notes: The value displayed for t-tests are the differences in the means across the groups. The value displayed for F-tests are the F-statistics. Standard errors are clustered at variable union. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.